

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**KLASIFIKASI SENTIMEN TERHADAP GUBERNUR DKI  
JAKARTA DI MEDIA SOSIAL *TWITTER* MENGGUNAKAN  
METODE *NAIVE BAYES CLASSIFIER***

**TUGAS AKHIR**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Serjana Teknik Pada Jurusan Teknik Informatika

Oleh:

**FARID FAHRIZA**  
**11351101165**



**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS NEGERI SULTAN SYARIF QASIM RIAU  
PEKANBARU**

**2021**

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**LEMBAR PERSETUJUAN**

**KLASIFIKASI SENTIMEN TERHADAP GUBERNUR DKI JAKARTA  
DI MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE *NAIVE  
BAYES CLASSIFIER***

**TUGAS AKHIR**

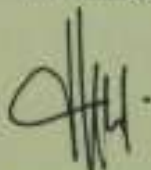
Oleh

**FARID FAHRIZA**

**11351101165**

Telah diperiksa dan disetujui sebagai laporan tugas akhir  
di Pekanbaru, pada tanggal 23 Februari 2021

Pembimbing I,



**Dr. Elin Haerani, ST, M.Kom**

**NIP. 19801010200710 1 002**

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**LEMBAR PENGESAHAN**

**KLASIFIKASI SENTIMEN TERHADAP GUBERNUR DKI JAKARTA DI  
MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES*  
CLASSIFIER**

**TUGAS AKHIR**  
Oleh  
**FARID FAHRIZA**  
11351101165

Telah dipertahankan di depan sidang dewan penguji  
sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika  
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau  
di Pekanbaru, pada tanggal 23 Februari 2021

Pekanbaru, 23 Februari 2021

Mengesahkan,

Dekan,



**Dr. Saiful Darmawati, M.Ag.**  
NIP. 19660604 199203 1 004

Ketua Jurusan,



**Dr. Elin Haerani, S.T., M.Kom.**  
NIP. 19810523 200710 2 003

**DEWAN PENGUJI**

Ketua	: Muhammad Irayad, ST, MT	
Pembimbing I	: Dr. Elin Haerani, ST, M.Kom	
Penguji I	: Surya Agustian, S.T, M.Kom.	
Penguji II	: Iis Alhanty, ST, M.Sc.	

**LEMBAR PENGESAHAN**



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# **KLASIFIKASI SENTIMEN TERHADAP GUBERNUR DKI JAKARTA DI MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER***

## **TUGAS AKHIR**

Oleh

**FARID FAHRIZA**  
**11351101165**

Telah dipertahankan di depan sidang dewan penguji  
sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika  
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau  
di Pekanbaru, pada tanggal 23 Februari 2021

Pekanbaru, 23 Februari 2021

Mengesahkan,

Dekan,

Ketua Jurusan,

**Dr. Ahmad Darmawi., M.Ag.**  
**NIP. 19660604 199203 1 004**

**Dr. Elin Haerani, S.T., M.Kom.**  
**NIP. 19810523 200710 2 003**

## **DEWAN PENGUJI**

Ketua	: Muhammad Irsyad, ST, MT	_____
Pembimbing I	: Dr. Elin Haerani, ST, M.Kom	_____
Penguji I	: Surya Agustian, S.T, M.Kom.	_____
Penguji II	: Iis Afrianty, ST, M.Sc.	_____

## LEMBAR HAK KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Pekanbaru, 23 Februari 2021

**FARID FAHRIZA**

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, Februari 2021  
Yang membuat pernyataan,

**Farid Fahriza**  
**11351101165**

UIN SUSKA RIAU



## LEMBAR PERSEMBAHAN



Maka nikmat tuhanmu yang manakah  
yang kamu dustakan? (QS. Ar-rahman 13)

Assalamu‘alaikum Warahmatullahi Wabarakatuh

Ku susun jari jemari ku diatas keyboard laptop ku sebagai pembuka kalimat persembahan. Diikuti dengan Bismillahirrahmanirrahim sebagai awal setiap memulai pekerjaan.

Sebuah langkah telah usai sudah, satu cita-cita telah tercapai. Sembah sujud serta puji dan syukurku pada-Mu Allah SWT. Tuhan semesta alam yang menciptakanku dengan bekal yang begitu teramat sempurna. Taburan cinta, kasih sayang, rahmat dan hidayat-Mu telah memberikan ku kekuatan, kesehatan,

semangat pantang menyerah dan memberkatiku dengan ilmu pengetahuan serta cinta yang pasti ada disetiap ummat-Mu. Atas karunia serta kemudahan yang Engkau berikan akhirnya tugas akhir ini dapat terselesaikan. Sholawat dan salam selalu ku limpahkan kepada Rasulullah Muhammad SAW.

Ku persembahkan tugas akhir ini untuk orang tercinta dan tersayang atas kasihnya yang berlimpah yaitu Ayahanda dan ibunda tercinta sebuah tulisan dari didikan kalian yang ku aplikasikan dengan ketikan hingga menjadi barisan tulisan

dengan beribu kesatuan, berjuta makna kehidupan, tidak bermaksud yang lain hanya ucapan TERIMA KASIH yang setulusnya tersirat dihati yang ingin ku sampaikan atas segala usaha dan jerih payah pengorbanan untuk anakmu selama ini. Hanya sebuah kado kecil yang dapat ku berikan dari bangku kuliahku yang memiliki sejuta makna, sejuta cerita, sejuta kenangan, pengorbanan, dan perjalanan untuk dapatkan masa depan yang ku inginkan atas restu dan dukungan yang kalian berikan.

Terimakasih untuk do‘a - do‘anya

Semoga tugas akhir ini bermanfaat bagi pembacanya Allahumma Aamiin

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# **KLASIFIKASI SENTIMEN TERHADAP GUBERNUR DKI JAKARTA DI MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER***

**FARID FAHRIZA**  
**11351101165**

Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Sultan Syarif Kasim Riau

## **ABSTRAK**

Twitter adalah salah satu jejaring sosial yang banyak diminati oleh penduduk dunia. Sebagian besar penduduk dunia menganggap bahwa Twitter adalah salah satu jejaring sosial yang mudah digunakan dan efisien. Mengkaji fenomena ini dapat dilihat juga bahwa penggunaan kode pagar atau hashtag (#) sebagai salah satu penanda dari fitur Twitter, ikut mempermudah sebuah topik untuk dibicarakan. Pengguna twitter bisa saling berkomunikasi dan saling membalas opini melalui tweet. Opini yang diberikan bisa bersifat positif dan negative. Dalam kasus ini di tujukan kepada Gubernur DKI Jakarta Anies Baswedan. Dari data komentar dan hastag (#) twitter Gubernur DKI Jakarta Anies Baswedan dapat diklasifikasikan sehingga didapat informasi baru. Dengan menggunakan teknik pengolahan data dalam *text mining*, Penelitian ini memanfaatkan hal tersebut dengan menggunakan metode *Naive Bayes Classifier*. Percobaan dilakukan pada 600 data komentar. Enam ratus data tersebut dibagi menjadi 2 kategori. Untuk mengetahui pengaruh jumlah data latih terhadap efektifitas *Naive Bayes Classifier* maka diambil beberapa kombinasi banyaknya dokumen latih dan dokumen uji. Secara berurutan kombinasi dokumen latih dan uji tersebut antara lain 90:10, 80:20, 70:30. Dengan menggunakan metode *Naive Bayes Classifier* didapatkan akurasi hasil klasifikasi berurut sebesar 65%, 70% dan 64%. Berdasarkan hasil pengujian yang telah dilakukan, akurasi tertinggi didapat pada pembagian data latih dan data uji 80:20 dengan nilai akurasi 70%. Aplikasi ini dapat dikembangkan untuk platform social media lainnya seperti facebook, instagram, path, kaskus, dan lainnya.

**Kata kunci : *Klasifikasi, Naive Bayes Classifier, Preprocessing, Text Mining, Twitter***



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# SENTIMENT CLASSIFICATION OF THE GOVERNOR OF DKI JAKARTA IN TWITTER SOCIAL MEDIA USING THE NAIVE BAYES CLASSIFIER METHOD

**FARID FAHRIZA**  
**11351101165**

Informatics Engineering  
Faculty of Science and Technology  
Sultan Syarif Kasim State Islamic University Riau

## ABSTRACT

Twitter is one of the social networks that is in great demand by the world's population. Most of the world's population consider that Twitter is one of the social networks that is easy to use and efficient. Examining this phenomenon can also be seen that the use of hashtag (#) code as a marker of the Twitter feature also makes it easier for a topic to be discussed. Twitter users can communicate with each other and reply to each other's opinions via tweets. Opinions given can be both positive and negative. In this case it was addressed to the Governor of DKI Jakarta Anies Baswedan. From the data of comments and hashtags (#) on twitter, the Governor of DKI Jakarta Anies Baswedan can be classified so that new information can be obtained. By using data processing techniques in text mining, this study makes use of this by using the Naive Bayes Classifier method. Experiments were carried out on 600 comment data. The six hundred data were divided into 2 categories. To determine the effect of the amount of training data on the effectiveness of the Naïve Bayes Classifier, several combinations of training documents and test documents were taken. Sequentially, the combination of practice and test documents includes 90:10, 80:20, 70:30. By using the Naïve Bayes Classifier method, the accuracy of the sequential classification results is 65%, 70% and 64%. Based on the results of tests that have been done, the highest accuracy is obtained in the distribution of training data and 80:20 test data with an accuracy value of 70%. This application can be developed for other social media platforms such as Facebook, Instagram, Path, Kaskus, and others.

***Keywords: Classification, Naive Bayes Classifier, Preprocessing, Text Mining, Twitter***

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## KATA PENGANTAR



*Assalamu'alaikum wa rahmatullahi wa barakatuh.*

*Alhamdulillah rabbil'alamin*, Segala puji hanya bagi Allah Subhanahu wata'ala, karena atas rahmat dan karunia-Nya penulis mampu menyelesaikan Laporan Tugas Akhir ini dengan lancar. Tidak lupa ucapan shalawat beriring salam untuk Baginda Rasulullah Muhammad Shalallahu'alaihi wa sallam, yang telah mengajarkan kita untuk menjadi manusia yang beradab dan beliau juga telah menjadi inspirator lahirnya zaman yang penuh dengan ilmu pengetahuan seperti saat ini. *Allahumma sholli'alaa sayyidina Muhammad wa'alaa ali sayyidina Muhammad.*

Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar kesarjanaan pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Banyak saran, pengalaman, pengetahuan, bimbingan dan dukungan menuju kebaikan yang penulis terima dari berbagai pihak hingga penulisan laporan ini dapat diselesaikan. Maka dari itu pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr.Suyitno, M.Ag, selaku Plt Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Ahmad Darmawi M,Ag, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Ibu Dr. Elin Haerani, ST., M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
4. Bapak Benny Sukma Negara, S.T, M.T, selaku Pembimbing Akademik penulis selama menjalani perkuliahan di Jurusan Teknik Informatika.

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5. Ibu Dr. Elin Haerani, ST., M.Kom, selaku Pembimbing I Tugas Akhir yang telah meluangkan waktu untuk membimbing penulis dan membagi ilmu, wawasan serta saran dan arahan sehingga penulis dapat memulai dan menyelesaikan penyusunan dan penulisan tugas akhir sampai selesai dengan baik.
6. Surya Agustian, S.T, M.Kom, selaku Penguji I Tugas Akhir yang telah memberikan masukan dan arahan kepada penulis dalam penyusunan tugas akhir ini.
7. Iis Afrianty, ST, M.Sc, selaku Penguji II Tugas Akhir yang telah memberikan masukan dan arahan kepada penulis dalam penyusunan tugas akhir ini.
8. Seluruh Bapak dan Ibu yang mengajar di jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau.
9. Kedua orang tua penulis yang selalu memberi semangat.
10. Teman seperjuangan angkatan 2013 yang telah memberikan dukungan moral maupun materi selama ini

Penulis menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat penulis harapkan untuk kesempurnaan laporan ini, yang dapat disampaikan ke alamat email penulis [fahrid.fahriza@students.uin-suska.ac.id](mailto:fahrid.fahriza@students.uin-suska.ac.id). Penulis berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya. Amin.

*Wassalamu'alaikum wa rahmatullahi wa barakatuh*

Pekanbaru, Desember 2020

Penulis



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## DAFTAR ISI

<b>ABSTRAK .....</b>	<b>ii</b>
<b>ABSTRACT .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>DAFTAR ISI.....</b>	<b>vi</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR PERSAMAAN.....</b>	<b>xii</b>
<b>DAFTAR SIMBOL .....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiv</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>I-1</b>
1.1 Latar Belakang .....	I-1
1.2 Rumusan Masalah .....	I-4
1.3 Batasan Masalah.....	I-4
1.4 Tujuan Penelitian.....	I-5
1.5 Sistematika penulisan .....	I-5
<b>BAB II LANDASAN TEORI .....</b>	<b>II-1</b>
2.1. Twitter.....	II-1
2.2. Text Mining .....	II-2
2.3. <i>Text Preprocessing</i> .....	II-2
2.4. <i>Term Frequency Inverse Document Frequency (TF-IDF)</i> .....	II-5
2.5. Naive Bayes Classifier.....	II-6
2.6. <i>White Box</i> .....	II-7
2.7. <i>Confusion Matrix</i> .....	II-7
2.8. Penelitian Terkait .....	II-8
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>III-1</b>
3.1. Identifikasi Masalah.....	III-2
3.2. Studi Pustaka.....	III-2
3.3. Pengumpulan Data .....	III-2
3.4. Analisa .....	III-2

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.4.1	Analisa Kebutuhan Data.....	III-3
3.4.2	Analisa Preprocessing .....	III-4
3.4.3	Analisa Pembobotan(TF-IDF) .....	III-5
3.4.4	Analisa Klasifikasi Naive Bayes.....	III-5
3.5.	Perancangan .....	III-6
3.6.	Implementasi dan Pengujian.....	III-6
3.7.	Kesimpulan dan Saran.....	III-7
<b>BAB IV ANALISA DAN PERANCANGAN .....</b>		<b>IV-1</b>
4.1.	Analisa Data.....	IV-1
4.1.1	Kebutuhan Data.....	IV-2
4.1.2	Pengumpul Data .....	IV-2
4.1.3	Text Preprocessing.....	IV-3
4.1.4	Tahap Pembobotan TF-IDF .....	IV-8
4.1.5	Proses Klasifikasi Metode Naive Bayes Classifier.....	IV-10
4.2.	Perancangan .....	IV-19
4.2.1	Perancangan Basis Data (Database) .....	IV-19
4.2.2	Perancangan Struktur Menu.....	IV-21
4.2.3	Perancangan Antarmuka (Interface) .....	IV-22
<b>BAB V IMPLEMENTASI DAN PENGUJIAN.....</b>		<b>V-1</b>
5.1.	Implementasi.....	V-1
5.1.1	Batasan Implementasi.....	V-1
5.1.2	Lingkungan Implementasi.....	V-1
5.1.3	Implementasi Aplikasi.....	V-2
5.2.	Pengujian .....	V-6
5.2.1	Pengujian White Box.....	V-6
5.2.2	Pengujian Confusion Matrix.....	V-12
5.2.3	Hasil Pengujian.....	V-16
<b>BAB VI PENUTUP .....</b>		
6.1	Kesimpulan .....	VI-1
6.2	Saran .....	VI-1
<b>DAFTAR PUSTAKA .....</b>		<b>XV</b>

# Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

## Gambar

## Halaman

3.1 Metodologi Penelitian .....	III-1
4.1 Umum Proses .....	IV-1
4.2 Flowchart Praproses Data .....	IV-4
4.3 Flowchart Pembobotan TF-IDF .....	IV-9
4.4 <i>Flowchart</i> Proses Klasifikasi .....	IV-10
4.5 Rancangan Struktur Menu.....	IV-21
4.6 Rancangan Antarmuka Halaman Beranda .....	IV-22
4.7 Rancangan Antarmuka Halaman Unduh Data Twitter .....	IV-23
4.8 Rancangan Halaman Data Twitter .....	IV-23
4.9 Rancangan Antar Muka Halaman Preproses.....	IV-24
4.10 Rancangan Antarmuka Halaman Data Latih.....	IV-24
4.11 Rancangan Antarmuka Halaman Data Uji .....	IV-25
5.1 Implementasi Antarmuka Halaman Beranda .....	V-2
5.2 Implementasi Antarmuka Halaman Unduh Data Twitter .....	V-3
5.3 Implementasi Antarmuka Halaman Data .....	V-3
5.4 Implementasi Halaman Pre-proses.....	V-4
5.5 Implementasi Antarmuka Halaman Data Latih.....	V-5
5.6 Implementasi Antarmuka Halaman Data Uji .....	V-5



## DAFTAR TABEL

Tabel	Halaman
2.1 Penelitian Terkait .....	II-8
4.1 Rincian <i>Dataset</i> .....	IV-2
4.2 Contoh Data Komentar .....	IV-3
4.3 Hasil <i>Cleaning</i> Data Komentar Anies baswedan .....	IV-4
4.4 Hasil <i>Case Folding</i> Data Komentar Anies baswedan .....	IV-5
4.5 Hasil <i>Tokenizing</i> Data Komentar Anies baswedan .....	IV-5
4.6 Hasil <i>Stemming</i> Data Komentar Anies baswedan .....	IV-6
4.7 Hasil <i>Normalisasi</i> Data Komentar Anies baswedan .....	IV-7
4.8 Hasil <i>Filtering</i> Data Komentar Anies baswedan .....	IV-7
4.9 Hasil <i>Negation Handling</i> Data Komentar Anies baswedan .....	IV-8
4.10 Hasil Pembobotan TF .....	IV-9
4.11 Hasil Pembobotan DF .....	IV-10
4.12 Hasil Pembobotan IDF .....	IV-10
4.13 Hasil Pembobotan TF-IDF .....	IV-11
4.14 Data <i>testing</i> (uji) .....	IV-17
4.15 Data Uji Setelah Praproses .....	IV-17
4.16 Contoh Tabel Data Latih Data Uji .....	IV-18
4.17 Nama Tabel Database Anies baswedan .....	IV-21
4.18 Struktur Tabel <i>Basicword</i> .....	IV-21
4.19 Struktur Tabel Kata Negasi .....	IV-22
4.20 Struktur Tabel Komentar .....	IV-22
4.21 Struktur Tabel Konfersi Negasi .....	IV-22
4.22 Struktur Tabel Konfigurasi .....	IV-23
4.23 Struktur Tabel Normalisasi .....	IV-23
4.24 dataset .....	IV-23
5.1 Pengujian White Box untuk Prioritas .....	V-6
5.2 Pengujian White Box untuk Frekuensi .....	V-7

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5.3 Pengujian White Box untuk Document Frequensy .....	V-9
5.4 Pengujian White Box untuk Inverse Document Frekuensi .....	V-9
5.5 Pengujian White Box untuk term frequensy inverse document frequency .....	V-10
5.6 Pengujian White Box untuk Probabilitas .....	V-11
5.7 Pengujian <i>White Box</i> untuk Probabilitas Uji .....	V-14
5.8 Pengujian Dataset 90% : 10% .....	V-14
5.9 Pengujian Dataset 80% : 20% .....	V-15
5.10 Pengujian Dataset 70% : 30% .....	V-15
5.10 Pengujian Dataset 70% : 30% .....	V-15
5.11 Jumlah Klasifikasi yang salah dan yang benar .....	V-15
5.12 Analisa Hasil Pengujian 90% : 10% .....	V-15

## DAFTAR PERSAMAAN

Persamaan	Halaman
(1) .....	II-6
(2) .....	II-7
(2.3) .....	II-8
(2.4) .....	II-8
(2.5) .....	II-8
(2.6) .....	II-9
(2.7) .....	II-10
(2.8) .....	II-10
(2.9) .....	II-10
(2.10) .....	II-10






### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



## DAFTAR SIMBOL

### Flowchart Diagram

Gambar	Keterangan
	Terminator: Simbol terminator (mulai/selesai) merupakan tanda bahwa sistem akan dijalankan atau berakhir.
	Proses: Simbol yang digunakan untuk melakukan pemrosesan data baik oleh user maupun komputer (sistem).
	Verifikasi: Simbol yang digunakan untuk memutuskan apakah valid atau tidak validnya suatu kejadian.
	Data: Simbol yang digunakan untuk mendeskripsikan data input/ output yang digunakan.
	Arus Data: Simbol yang digunakan untuk menggambarkan arus data di dalam sistem.

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Media sosial merupakan situs dimana seseorang dapat membuat web page pribadi dan terhubung dengan setiap orang yang tergabung dalam media sosial yang sama untuk berbagi informasi dan berkomunikasi. Tidak dapat dipungkiri bahwa media sosial mempunyai pengaruh yang besar dalam kehidupan seseorang. Seseorang yang asalnya kecil bisa menjadi besar dengan media sosial, begitu pula sebaliknya (Basri, 2017). Dari sekian banyak media sosial yang paling cepat pertumbuhannya adalah twitter. Twitter merupakan microblogging, layanan pesan pendek dengan kemampuan 140 karakter dalam bentuk text yang cukup populer digunakan saat ini, bahkan Indonesia telah dikenal sebagai pengguna Twitter terbesar di dunia, selain Facebook. Jumlah pengguna harian Twitter secara global diklaim meningkat. Pada laporan finansial Twitter kuartal ke-3 tahun 2019, pengguna aktif harian di platform Twitter dicatat meningkat 17 persen, ke angka 145 juta pengguna. Indonesia diklaim menjadi salah satu negara yang pertumbuhan pengguna aktif harian Twitter-nya paling besar. Klaim ini disampaikan oleh Dwi Adriansah selaku Country Industry Head Twitter. Twitter tidak hanya di gunakan oleh para remaja tetapi, juga digunakan oleh tokoh politik salah satu adalah Gubernur DKI Jakarta Anies Baswedan.

Gubernur DKI Jakarta Anies Baswedan merupakan salah satu tokoh dari dunia pendidikan dan politik yang memiliki peranan penting di Indonesia. Saat ini beliau menjabat sebagai gubernur DKI Jakarta sejak tahun 2017. Akan tetapi sebelum itu Gubernur DKI Jakarta Anies Baswedan pernah menjabat sebagai menteri pendidikan dan kebudayaan di Indonesia. Dengan berbagai pengalaman di bidang pendidikan, beliau berhasil mendirikan gerakan Indonesia Mengajar pada tahun 2010 dan masih aktif sampai sekarang. (Fifiarini, Halim, Apriantour, & Nannisa, 2019). Walaupun Gubernur DKI Jakarta Anies Baswedan telah

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

memperoleh banyak prestasi, namun ada juga yang memiliki sentimen buruk terhadap kinerja Gubernur DKI Jakarta Anies Baswedan.

Analisa sentimen atau bisa disebut juga dengan opinion mining merupakan bidang studi yang menganalisa opini, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap suatu entitas seperti produk, pelayanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya. Dalam komentar maupun hastag (#) yang di tujukan kepada Gubernur DKI Jakarta Anies Baswedan terdapat banyak sentimen negatif dan positif. Ini dapat dijadikan sumber data dalam penelitian text mining. Dari data komentar dan hastag (#) twitter Gubernur DKI Jakarta Anies Baswedan dapat diklasifikasikan sehingga didapat informasi baru.

Informasi baru tersebut dapat diklasifikasikan untuk memperoleh nilai probabilistik sederhana yang dihitung dari sekumpulan probabilitas yang menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan dengan menggunakan teorema Naive Bayes. Algoritma Naive Bayes ini mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan. (Alfa, 2010).

Beberapa penelitian yang terkait mengenai analisis sentiment adalah, “Analisis Sentimen Menggunakan *Naive Bayes*” melibatkan 200 data ulasan, dan dilakukan perbandingan penggunaan data latih 70% dan data uji 30% dengan teks pre-processing menghasilkan nilai akurasi lebih unggul sebesar 96,9% (Gunawan, Fauzi, & Adikara, 2017). Selanjutnya penelitian terkait yang menggunakan berbagai metode teks klasifikasi antara lain, “Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan *Naive Bayes*” dari hasil penelitian ini



menggunakan tiga kelas maka dapat menghasilkan nilai akurasi sebesar 93%, dan tingkat precision positif sebesar 90%, precision negatif sebesar 90% dan sentimen yang netral 100% (Rustiana & Rahayu, 2017). Selanjutnya penelitian terkait yang berjudul “Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier” dari hasil penelitian ini algoritma *Naive Bayes Classifier* memberikan nilai akurasi sebesar 97%. Untuk nilai akurasi tiap sentimennya yaitu 96% untuk sentimen negatif dan 98% untuk sentimen positif (Mahardhika & Zuliarso, 2018). Selanjutnya penelitian terkait yang berjudul “Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode *Naive Bayes*” dari hasil penelitian ini hasil yang didapatkan dari akurasi *Naive Bayes* memperoleh ketepatan 80% (Nugroho, Chrisnanto, & Wahana, 2016).

Dari penelitian yang telah dilakukan disimpulkan bahwa analisa sentimen menggunakan metode klasifikasi *Naive Bayes Classifier* mendapatkan akurasi yang tinggi, maka dari itu penulis memutuskan untuk memilih metode ini sebagai metode penelitian yang sedang dilakukan. Dari penjelasan latar belakang diatas maka penulis melakukan penelitian tugas akhir tentang “Klasifikasi Sentimen Terhadap Gubernur DKI Jakarta Anies Baswedan di Media Sosial Twitter Menggunakan Metode *Naive Bayes Classifier*”.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah di uraikan diatas yang menjadi pokok permasalahan dalam hal ini adalah : “Bagaimana menerapkan metode Klasifikasi Sentimen Terhadap Gubernur DKI Jakarta Anies Baswedan di Media Sosial *Twitter* Menggunakan Metode *Naive Bayes Classifier*”.

## 1.3. Batasan Masalah

Agar permasalahan yang dibahas tidak keluar dari pokok permasalahan yang telah dirumuskan, maka diberikan beberapa batasan masalah antara lain :

1. Data ulasan yang digunakan adalah sebanyak 600 data
2. Klasifikasi terdiri dari 2 kelas yaitu positif dan negatif.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. Bahasa yang digunakan Bahasa Indonesia.

## 1.4. Tujuan Penelitian

Adapun tujuan yang dilakukan dalam penelitian Tugas Akhir ini yaitu:

1. Mengklasifikasikan sentiment terhadap Gubernur DKI Jakarta Anies Baswedan yang terdiri dari positif, dan negatif dengan proses text mining menggunakan metode klasifikasi *Naive Bayes Classifier*.
2. Menguji akurasi metode *Naive Bayes Classifier* dalam Pengklasifikasian Sentimen terhadap Gubernur DKI Jakarta Anies Baswedan berdasarkan ulasan di Twitter.

## 1.5. Sistematika Penulisan

Sistematika penulisan dalam penelitian tugas akhir ini terdiri dari 6 bab, antara lain:

### BAB I PENDAHULUAN

Pada bab ini berisi latar belakang, batasan masalah, rumusan masalah, tujuan dan sistematika penulisan.

### BAB II LANDASAN TEORI

Bab ini menjelaskan teori singkat tentang hal-hal yang berhubungan dengan penelitian yang diteliti yakni klasifikasi sentimen terhadap Raditya Dika pada komentar video Raditya Dika di Twitter menggunakan metode *Naive Bayes Classifier*.

### BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang beberapa rangkain tahapan dalam pembuatan aplikasi, mulai dari pengumpulan data, analisa dan perancangan, tahap implementasi dan pengujian yang digunakan, hingga kesimpulan dan saran.

### BAB IV ANALISA DAN PERANCANGAN

Bab ini berisi tentang analisa kebutuhan, kemudian melakukan perancangan terhadap desain yang dibutuhkan untuk merancang dan

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

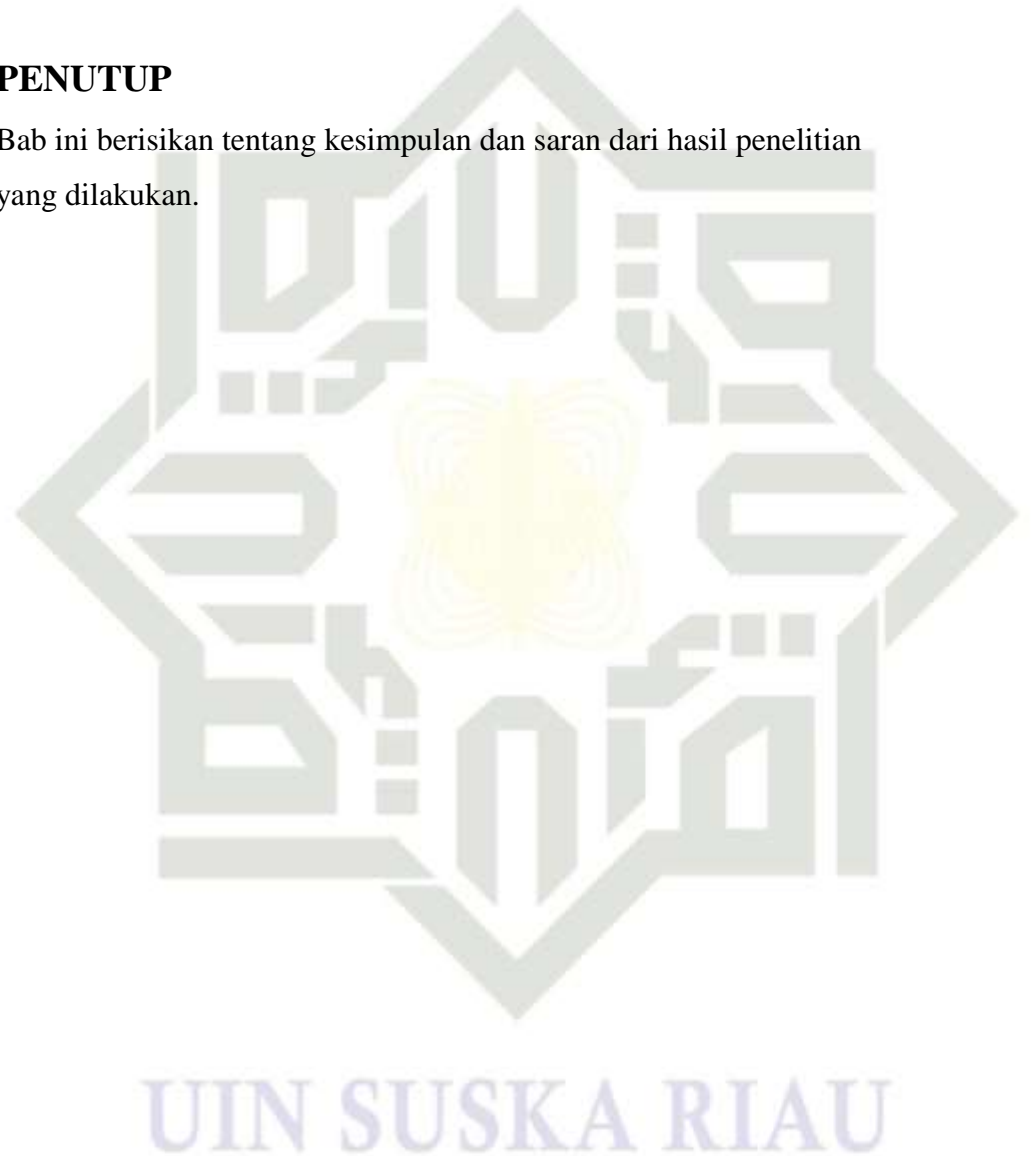
membangun aplikasi klasifikasi sentimen pada komentar video *Anies baswedan* menggunakan metode naive bayes *classifier*.

## **IMPLEMENTASI DAN PENGUJIAN**

Bagian bab ini berisi batasan dari implementasi, bentuk hasil aplikasi yang telah di rancang dan melakukan proses pengujian.

## **PENUTUP**

Bab ini berisikan tentang kesimpulan dan saran dari hasil penelitian yang dilakukan.





Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## BAB II

### LANDASAN TEORI

#### 2.1. Twitter

Twitter merupakan sebuah situs web yang dimiliki dan dioperasikan oleh Twitter, inc. Media sosial ini juga merupakan salah satu layanan jejaring sosial dan microblog daring yang memungkinkan para penggunanya untuk mengirim, menerima dan membaca pesan berbasis teks yang jumlah karakternya mencapai 140 karakter, yang dikenal dengan sebutan kicauan (tweet). Twitter adalah salah satu jejaring sosial yang banyak diminati oleh penduduk dunia. Sebagian besar penduduk dunia menganggap bahwa Twitter adalah salah satu jejaring sosial yang mudah digunakan dan efisien. Maka dari itu, pengguna Twitter dari tahun ke tahun selalu meningkat drastis. Mengkaji fenomena ini dapat dilihat juga bahwa penggunaan kode pagar atau hashtag (#) sebagai salah satu penanda dari fitur Twitter, ikut mempermudah sebuah topik untuk dibicarakan. Fitur ini biasanya digunakan untuk membedakan antara sebuah topik dengan topik lain yang memiliki kalimat yang mirip atau serupa di dalam aktivitas di media sosial Twitter. Penggunaan tanda pagar (hashtag) di Twitter pada negara-negara lain tujuannya agar orang lain paham tentang topik yang dimaksudkan oleh pengguna hashtag. Namun kenyataan yang lain terjadi di Indonesia, di mana tujuan penggunaan tanda pagar tersebut salah satunya adalah untuk menaikkan popularitas sebuah topik tertentu. Dengan penggunaan hashtag ini pengguna lainnya ikut meretwit dan juga membuat hash tag yang sama. Hal ini untuk meningkatkan intensitas tentang topik yang dimaksud. (Juditha, 2015).

Twitter juga merupakan media sosial yang banyak mendapatkan perhatian masyarakat Indonesia. Secara global, berdasarkan data pada Desember 2014, terdapat 284 juta pengguna aktif layanan tersebut. Jumlah pengguna Twitter di Indonesia sudah mencapai angka 50 juta. Jumlah ini diprediksi makin terus bertambah. Hal ini menyebabkan Indonesia merupakan pasar yang paling menguntungkan (Deliusno, 2015).

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Terdapat beberapa istilah – istilah yang umum ditemui pada Twitter, yaitu:

1. *Timeline* yaitu daftar tweet terbaru dari pengguna twitter yang diikuti pemilik akun, termasuk tweet yang dibuat pemilik akun.
2. *Direct Message* yaitu fasilitas berkirim pesan antar pengguna secara lebih *private*.
3. *Trending Topic* daftar tema yang lagi hangat diperbincangkan di kalangan pengguna twitter.
4. *Tweet* merupakan informasi yang terdiri dari pesan 140 karakter. Tweet berisi berita terbaru yang berkaitan dengan hal-hal yang pemilik akun sukai.
5. *Reply tweet* atau *response tweet* yaitu komentar atau balasan atas tweet.
6. *Retweet* adalah menyalin seluruh isi tweet dari akun lain.
7. *Follow* adalah mengikuti akun dan informasi yang disampaikan oleh pengguna.
8. *Follower* adalah pengikut yang mengikuti akun orang lain.
9. *Mention (@)* untuk menyebut *username* akun yang diajak berkomunikasi.
10. *Hastags (#)* digunakan untuk mendandai kata kunci topic diskusi atau informasi yang dibagikan agar mudah dicari.

## 2.2. Text Mining

Text mining merupakan bagian penerapan dari data mining yang digunakan untuk mencari pola dalam teks. Berbeda dengan data mining yang menggunakan data biasa, dalam text mining ini data yang digunakan adalah data dalam bentuk dokumen atau teks (Feizar, 2014). Untuk mempermudah mengetahui jenis konten dari sejumlah data tweet, maka perlu dilakukan proses Text Mining terhadap data tweet tersebut dengan menerapkan teknik clustering (Zhao, 2030).

Pada Text Mining, teknik clustering digunakan untuk mengelompokkan data tekstual berdasarkan kesamaan konten yang dimiliki ke dalam beberapa klaster, sehingga didalam setiap klaster akan berisi data tekstual dengan konten semirip mungkin (Feldman & Sanger, 2007).

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 2.3. Text Preprocessing

*Text preprocessing* adalah proses awal dalam *text mining*. *Preprocessing* dilakukan untuk menghilangkan bagian teks yang tidak diperlukan sehingga mendapatkan data yang berkualitas untuk dieksekusi (Sanjaya dan Absar, 2015).

Adapun tahapan preprocessing berdasarkan proses yaitu:

### 1. *Cleaning*

*Cleaning* merupakan proses membersihkan dokumen dari komponen-komponen yang tidak memiliki hubungan dengan informasi yang ada pada dokumen. Seperti Karakter atau simbol (@#\$%^&\*()\_+”:{}<.,?!~/[]), angka, link url (<http://link.com>), hashtag (#), dan mention @username.

### 2. *Case Folding*

*Case folding* tahap mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter (pembatas).

### 3. *Tokenizing*

*Tokenizing/parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Dan spasi digunakan untuk memisahkan antar kata yang disebutkan.

### 4. *Normalisasi*

Merupakan proses mengembalikan kata-kata yang tidak baku kedalam bahasa baku sesuai dalam kamus bahasa Indonesia.

### 5. *Negation Handling*

*Negation Handling* berfungsi untuk meningkatkan akurasi dalam klasifikasi, dimana apabila terdapat kata seperti tidak, bukan dan lainnya, hal ini akan membuat makna kata berikutnya menjadi berlawanan dengan makna sesungguhnya. Contohnya seperti kata "tidak baik" menjadi "buruk".

Berikut ini beberapa langkah-langkah *negation handling* menurut Vivek Narayanan (Narayanan, Arora and Bhatia, 2030).

1. *Inisialisasi negasi = false*
2. *For each kata in dok*



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

*Ifnegasi = True*

Ubah kata menjadi = “kata\_negasi” + “kata”

*If kata = “tidak” or “bukan” or “nggak” or “belum”*

*Negasi = True*

*Elsenegasi = false*

Proses *negation handling* yang diatas adalah mencari kata yang bersifat negasi, jika ditemukan kata negasi maka kata negasi tersebut digabung dengan kata selanjutnya dan selanjutnya digantikan dengan lawan kata setelah kata negasi tersebut.

#### 6. Filtering

*Filtering* adalah tahap mengambil kata-kata penting dari hasil *tokenizing*. Proses *filtering* dapat menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/ stopworld* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*.

#### 7. Stemming

Stemming merupakan proses yang digunakan untuk menemukan kata dasar dari sebuah kata dengan menghilangkan semua imbuhan (*affixes*) baik yang merupakan awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*), dan kombinasi dari awalan serta akhiran (*confixes*) pada kata turunan.

Algoritma *stemming* kata pada bahasa Indonesia dengan performa yang paling baik saat ini (memiliki jenis kesalahan *stemming* yang paling sedikit) adalah Algoritma *Enhanced Confix Stripping* (ECS) (Arifin, A. Z., Mahendra, I. P., & Ciptaningtyas, 2009). Perbaikan yang dilakukan oleh stemming ECS adalah perbaikan beberapa aturan pada tabel acuan pemenggalan imbuhan. Selain itu, algoritma stemming ECS juga menambahkan langkah pengembalian akhiran jika terjadi penghilangan akhiran yang seharusnya tidak dilakukan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 2.4. Term Frequency Inverse Document Frequency (TF-IDF)

*Term Frequency Inverse Document Frequency* atau TF-IDF adalah sebuah metode yang biasanya digunakan untuk menetapkan nilai frekuensi disetiap kata dalam dokumen atau artikel. Dalam perhitungan setiap kata seberapa relevan kata tersebut dalam sebuah dokumen (Hendy Evan & Sigit Purnomo, 2014).

*Term Frequency Inverse Document Frequency* atau TF-IDF merupakan algoritma yang pada umumnya digunakan dalam pengolahan data yang besar, algoritma ini untuk pembobotan dalam setiap kata dan kategori untuk mencari kesamaan kata kunci pada katagori yang ada. Sebelum proses pembobotan kata akan dilakukan tahapan-tahapan *text Pre-processing* berikut :

1. *case folding*
2. *tokenizing*
3. *filtering*
4. *stemming*

kemudian selanjutnya adalah proses penghitungan bobot TF-IDF, bobot *query relevance* dan bobot *similarity* (Marlinda, Lina & Rianto, 2030)

Untuk menghitung nilai TF-IDF yaitu:

$$w_{ij} = tf_{ij} \times \log(D/df_j) \quad (2.1)$$

Keterangan:

$w_{ij}$  : bobot *term* terhadap dokumen

$tf_{ij}$  : jumlah kemunculan *term* dalam dokumen

$D$  : total jumlah dokumen

$df_j$  : jumlah dokumen yang terdapat *term*

Berikut merupakan cara menghitung nilai dari TF-IDF:

1. Menghitung nilai *tf*
2. Menghitung nilai *inverst document frequency*
3. Menghitung nilai *tf* dan *idf*

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 2.5. Naive Bayes Classifier

Naive Bayes *Classifier* adalah algoritma untuk mencari probabilitas dengan nilai tertinggi untuk pengklasifikasian data uji untuk kategori yang tepat. pada klasifikasi data teks memiliki dua tahapan yaitu data *training* (pelatihan) kepada dokumen yang sudah diketahui kategorinya, dan tahap kedua yaitu data *testing* (pengujian), adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Tahapan dari algoritma *Naive Bayes Classifier* adalah :

1. Menghitung jumlah kelas / label
2. Menghitung jumlah kasus per kelas
3. Kalikan semua variable kelas
4. Bandingkan hasil per kelas

Pada metode *Naive Bayes Classifier* dokumen akan dipresesntasikan. Saat pengklasifikasi maka algoritma akan mencari probabilitas dengan nilai tertinggi pada semua kategori dokumen yang diuji (Jayanti et al, 2016), yang persamaannya adalah :

$$P(C|X) = \frac{p(x|c)p(c)}{p(x)} \quad (2.2)$$

Keterangan :

- X : Data dengan class yang belum diketahui  
 C : Hipotesis data merupakan suatu class spesifik  
 P(c|x) : Probabilitas hipotesis berdasar kondisi (posteriori probability)  
 P(c) : Probabilitas hipotesis (prior probability)  
 P(x|c) : Probabilitas berdasarkan kondisi pada hipotesis  
 P(x) : Probabilitas c

Rumus diatas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali dikenal dengan istilah prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik sampel secara global (disebut juga evidence). Karena itu, rumus diatas dapat pula ditulis sebaga berikut:



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$Posterior = \frac{prior \times likelihood}{evidence} \quad (2.3)$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai nilai posterior kelas lainnya untuk ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $(c|x_1, \dots, x_n)$  menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|X_1, \dots, X_n) &= P(C)P(X_1, \dots, X_n|C) \\ &= P(C)P(X_1|c)(X_2, \dots, X_n|C, X_1) \\ &= P(C)P(X_1|c)P(X_2|C, X_1)(X_3, \dots, X_n|C, X_1, X_2) \\ &= P(C)P(X_1|c)P(X_2|C, X_1)P(X_3|C, X_1, X_2) \\ &\quad \dots P(X_n|C, X_1, X_2, \dots X_{n-1}). \end{aligned} \quad (2.4)$$

Ide dasar dari aturan *Naive Bayes Classifier* hasil dari hipotesis atau peristiwa dapat diperkirakan berdasarkan pada beberapa bukti yang diamati. Ada beberapa hal penting dari aturan *Naive Bayes Classifier* yaitu :

1. Sebuah probabilitas awal/prior adalah probabilitas dari suatu hipotesa sebelum bukti diamati.
2. Sebuah probabilitas akhir adalah probabilitas dari suatu hipotesis setelah bukti diamati.

## 2.6. White Box

*White box* atau *Glass Box testing* adalah pengujian yang didasarkan pada pengecekan terhadap detail perancangan, menggunakan struktur kontrol dari desain program secara prosedural untuk membagi pengujian kedalam beberapa *Test Case* (Nidra & Dondeti, 2012). Penentuan kasus uji disesuaikan dengan struktur aplikasi, pengetahuan mengenai program digunakan untuk mengidentifikasi kasus uji tambahan.

Tujuan penggunaan *white box* untuk menguji semua statement program. Penggunaan metode pengujian *white box* dilakukan untuk :

1. Memberikan jaminan bahwa semua jalur independen suatu modul digunakan minimal satu kali.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Menggunakan semua keputusan logis untuk semua kondisi *true* atau *false*.
3. Mengeksekusi semua perulangan pada batasan nilai dan operasional pada setiap kondisi.
4. Menggunakan struktur data internal untuk menjamin validitas jalur keputusan.

## 2.7. Confusion Matrix

*Confusion matrix* adalah salah satu metode yang dapat dipakai untuk menghitung kinerja pada suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Prasetyo, 2012).

Pada tahap ini akan mengevaluasi hasil percobaan, membandingkan dan menganalisis terhadap kinerja klasifikasi teks. Banyak cara pengukuran yang telah digunakan, seperti *Precision* dan *Recall*, *error*, akurasi, dan lainnya. Pada penelitian ini metode evaluasi yang digunakan adalah *Recall*, *Precision*, *Accuracy*, *Support*, dan *F1 Score*

Berikut ini adalah perhitungan untuk memperoleh nilai akurasi.

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \quad (2.6)$$

Keterangan :

TP = *True positives*, adalah jumlah data kelas positif yang diklasifikasikan dengan benar oleh sistem.

TN = *True negatives*, adalah jumlah data kelas negatif yang diklasifikasikan dengan benar oleh sistem .

FN = *False negatives*, adalah jumlah data kelas negatif namun diklasifikasikan positif.

FP = *False positives*, merupakan jumlah data kelas positif namun diklasifikasikan negatif.

Berikut ini adalah perhitungan untuk memperoleh nilai *precision*.

$$Precision = (TP) / (TP + FP) \quad (2.7)$$

Keterangan :

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

TP = *True positives*, adalah jumlah data kelas positif yang diklasifikasikan dengan benar oleh sistem.

FP = *False positives*, merupakan jumlah data kelas positif namun diklasifikasikan negatif.

Berikut ini adalah perhitungan untuk memperoleh nilai *recall*.

$$Recall = (TP) / (TP + FN) \quad (2.8)$$

Keterangan :

TP = *True positives*, adalah jumlah data kelas positif yang diklasifikasikan dengan benar oleh sistem.

FN = *False negatives*, adalah jumlah data kelas negatif namun diklasifikasikan positif.

Berikut ini adalah perhitungan untuk memperoleh nilai *support*.

$$Support = (TN) / (TN + FP) \quad (2.9)$$

Keterangan :

TN = *True negatives*, adalah jumlah data kelas negatif yang diklasifikasikan dengan benar oleh sistem .

FP = *False positives*, merupakan jumlah data kelas positif namun diklasifikasikan negatif.

Dan berikut ini adalah perhitungan untuk memperoleh nilai *support*.

$$F1 \text{ Score} = 2 * (Recall * Precision) / (Recall + Precision) \quad (2.10)$$

## 2.8. Penelitian Terkait

Pada tabel 2.1 berikut ini menjelaskan tentang penelitian terkait yang telah dilakukan sebelumnya.

No	Nama Penulis	Judul	Tahun	Metode	Hasil
----	--------------	-------	-------	--------	-------



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	Teguh Ansyor Lorosae, Burhanudin Dwi Prakoso, Saifudin, Kusrini	Analisis sentimen berdasarkan opini masyarakat pada twitter menggunakan naive bayes	2018	<i>Naive bayes classsifier</i>	Dari hasil penelitian didapatkan tingkat akurasi tertinggi 84%
	Burhanudin, Yuniarti Musa'adah, Yaya Wihardi	Klasifikasi komentar spam pada Twitter menggunakan metode naive bayes, svm dan k-nn	2018	<i>Naive bayes classsifier ,SVM, K- NN</i>	Multinomial naive bayes, bernouli naive bayes, nu SVC, dan linear svc memiliki tingkat precision, recall dan fi- scores paling tinggi dibanding metode lainnya
3	Sunardi, Abdul Fadlil, Suprianto	Analisis Sentimen Menggunakan Metode <i>Naive Bayes Classifier</i> Pada Angket Mahasiswa	2018	<i>Naive bayes classsifier</i>	Hasil dari klasifikasi metode pengujian ini memiliki tingkat akurasi 80%
4	Haditsah Annur	Klasifikasi Masyarakat Miskin Menggunakan Metode <i>Naive Bayes</i>	2018	<i>Naive Bayes Classifier</i>	Berdasarkan penelitian yang dilakukan dihasilkan kesimpulan bahwa, Sistem klasifikasi masyarakat miskin di wilayah pemerintahan Kecamatan Tibawa Kab. Gorontalo dapat direkayasa dan Berdasarkan hasil pengujian confussion

# **Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

					matrix dengan teknik split validasi, penggunaan metode klasifikasi <i>Naive Bayes</i> terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori Good. Sementara nilai Precision sebesar 92% dan Recall sebesar 86%.
5	Agnes Rossi Trisna Lestari, Rizal Setya Perdana, M. Ali Fauzi	Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji	2017	<i>Naive Bayes Classifier</i>	Dari hasil pengujian akurasi, diperoleh 68,52% untuk kondisi pembobotan tekstual, 75,93% untuk pembobotan non-tesktual, dan 74,81% untuk kondisi penggabungan dengan nilai konstanta 0,5 untuk tekstual dan 0,5 untuk non-tekstual.
	Muhammad	Klasifikasi	2017	<i>Naive</i>	Pada penelitian

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Abdurokhim	pengaduan pelayanan publik dikota semarang dari data twitter menggunakan metode naive bayes classifier		<i>bayes classsifier</i>	ini menghasilkan akurasi menggunakan naive bayes yaitu sebesar 77%
Effrida Manalu, Fricles Ariwisanto Sianturi, mamed Rofendy Manalu	Penerapan algoritma naive bayes untuk memproduksi jumlah produksi barang berdasarkan data persediaan dan jumlah pemesanan pada cv.papa dan mama pastries	2017	<i>Naive bayes classsifier</i>	Memberikan solusi dalam memprediksi persediaan roti dan penerimaan pesanan roti yang di ambil berdasarkan data pada tahun lalu, sehinggadapat membantu dalam menentukan stok roti sesuai dengan kebutuhan dan keinginan serta menghasilkan alternatif pilihan stok roti
Evasaria.M, Sipayung, Herastia maharani, Ivan Zafanya	Perancangan sistem analisis sentimen komentar pelanggan menggunakan metode naive bayes classifier	2016	<i>Naive bayes classsifier</i>	Tingkat akurasi dalam penentuan kategori adalah sebesar 77,14% dan 75,42% dalam penentuan sentimen memiliki tingkat precission 99,12% dan recall 72,9%



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

	Ahmad Fathan Hidayatullah, Azhari SN	Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter	2014	<i>Naive bayes classsifier</i>	Perhitungan akurasi klasifikasi menggunakan metode naive bayes classsifier dengan fitur term frequency diperoleh sebesar 73,81%
10	Yuda Septian Nugroho	Data Mining Menggunakan Algoritma <i>Naive Bayes</i> Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro	2014	<i>Naive bayes classsifier</i>	Perolehan hasil dari penggunaan metode <i>Naive Bayes</i> yaitu akurasi kelulusan sebesar 82,08%
11	Faishol Nurhuda, Sari Widya Sihwi, Afrizal Doewes	Analisis sentimen masyarakat terhadap calon presiden indonesia 2014 berdasarkan opini dari twitter menggunakan metode naive bayes classsifier	2014	<i>Naive bayes classsifier</i>	Pasangan Prabowo Subianto-Hatta Rajasa lebih unggul sebesar 53% dibandingkan percakapan terhadap pasangan jokowi yang mendapat 47%
12	Wayan Firdaus Mahmudy, Agus Wahyu Widodo	Klasifikasi artikel berita secara otomatis menggunakan metode naive bayes classsifier yang dimodifikasi	2014	<i>Naive bayes classsifier</i>	Dengan menggunakan metode naive bayes didapatkan hasil klasifikasi berturut-turut 54%, 65%, 69%, 71%, dan 76%

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

30	Sigit Suryono, Ema Utami, Taufiq Luthfi	Klasifikasi sentimen pada twitter dengan menggunakan naive bayes classifier	2014	<i>Naive bayes classsifier</i>	Hasil pembahasan dapat di simpulkan bahwa tingkat akurasi sebesar 66,76%
14	Akhmad Pandu Wijaya, Heru Agus Santoso	Naive bayes classifier pada klasifikasi dokumen untuk identifikasi konten E- government	2014	<i>Naive bayes classsifier</i>	Hasil dari klasifikasi dokumen menggunakan naive bayes classifier sebesar 85%

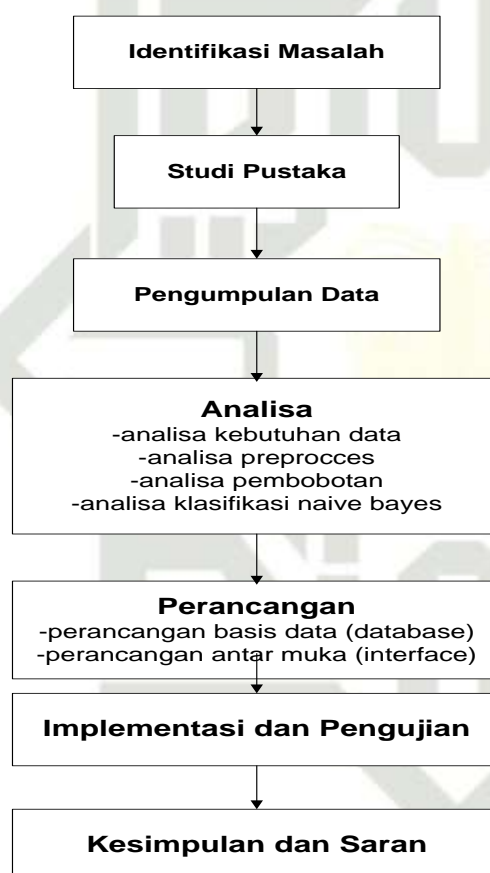
**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## BAB III

### METODOLOGI PENELITIAN

Metodologi penelitian menjelaskan tentang langkah-langkah yang ditempuh saat melakukan penelitian untuk menjawab pertanyaan dari rumusan masalah pada penelitian. Adapaun langkah-langkah penelitian dapat dilihat pada gambar 3.1 sebagai berikut.



**Gambar 3.1 Metodologi Penelitian**



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### **3.1. Identifikasi Masalah**

Tahap ini merupakan tahapan awal dari metodologi penelitian untuk menentukan latar belakang dan rumusan masalah serta tujuan dari penelitian ini. Selain itu pada tahap ini juga menentukan batasan-batasan penelitian. Adapun hal-hal yang dilakukan pada tahap identifikasi masalah yaitu sebagai berikut:

#### **1. Latar belakang**

Merupakan hal dasar yang membuat penelitian menjadi perlu dilakukan, dalam penelitian ini fenomena Gubernur DKI Jakarta Anies Baswedan di Media Sosial *Twitter* dan kecenderungan penggunaanya untuk menyampaikan komentar (ulasan) terhadap Gubernur DKI Jakarta Anies Baswedan melalui *Twitter* menjadi salah satu latar belakang penelitian.

#### **2. Tujuan Penelitian**

Merupakan hal-hal yang ingin dicapai dalam penelitian. Dalam penelitian ini tujuan utamanya yaitu untuk mengklasifikasikan Analisis Sentimen kedalam kelas positif atau negatif dengan tingkat akurasi yang tinggi.

#### **3. Rumusan Masalah**

Rumusan masalah adalah masalah-masalah yang harus di selesaikan dalam penelitian, contohnya metode apa yang akan digunakan, dan bagaimana memperoleh data komentar (ulasan) dari *Twitter*.

#### **4. Batasan Masalah**

Batasan masalah menjadi salah satu yang juga harus dilakukan pada tahap identifikasi agar penelitian menjadi lebih fokus sehingga tujuan penelitian dapat tercapai.

### **3.2. Studi Pustaka**

Pada tahap pengumpulan data ini yang dilakukan adalah pengumpulan data dan segala informasi yang meliputi pencarian referensi berupa jurnal-jurnal, buku yang berupa *text book* maupun *e-book*, media *online* dan referensi-referensi lainnya yang berhubungan dengan penelitian ini. Referensi yang dikumpulkan adalah referensi yang berkaitan dengan teori-teori mengenai penelitian serupa yang sebelumnya telah dilakukan.

### 3.3. Pengumpulan Data

Pada penelitian ini, pengumpulan data komentar *twitter* dengan cara menyeleksi *tweet* berdasarkan *hashtag* (#). lalu menyalinnya ke dalam *Microsoft Excel*. Dan langsung melakukan pelabelan manual.

### 3.4. Analisa

Setelah mengidentifikasi masalah dan telah melakukan pengumpulan data, selanjutnya tahapannya adalah analisa. Analisa merupakan sebuah tahapan yang didalamnya mempelajari serta mengevaluasi suatu permasalahan dan kasus yang terjadi tujuannya untuk mengetahui gambaran jelas mengenai penelitian yang dilakukan. Adapun tahapan analisa adalah sebagai berikut:

#### 3.4.1 Analisa Kebutuhan Data

Analisa kebutuhan data akan membahas mengenai kategori data yang akan digunakan dalam proses analisa serta bagaimana cara dalam pengumpulan datanya, dengan rincian sebagai berikut:

##### 1. Kategori Data

Dalam penelitian ini, data yang dibutuhkan dibagi menjadi 2 kategori (class) komentar yaitu kategori positif dan kategori negatif.

- Kategori positif yaitu komentar (ulasan) pengguna yang dinyatakan pengguna suka dengan Gubernur DKI Jakarta Anies Baswedan.
- Kategori negatif yaitu pengguna yang tidak suka dengan Gubernur DKI Jakarta Anies Baswedan.

##### 2. Pembagian Data

Pada penelitian ini, data komentar dibagi menjadi 2 yaitu data *latih* dan data *uji*. Data *latih* merupakan kumpulan data komentar yang akan dijadikan sebagai data *latih* pada sistem. Pada penelitian ini data yang terkumpul dilakukan pengkategorian secara manual. Sedangkan data *uji* merupakan data komentar baru yang belum memiliki kategori yang mana akan digunakan sebagai analisa dan uji oleh sistem berdasarkan data *latih* yang ada. Ada beberapa kombinasi yang baik digunakan untuk pembagian data yaitu 70% : 30%, dengan 80% : 20% dan 90% : 10%.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3.4.2 Analisa Preprocessing

Pada tahapan ini memaparkan bagaimana teks akan dipersiapkan menjadi data yang diproses lebih lanjut. Adapun tahapannya adalah:

1. pertama melakukan proses *cleaning* yaitu membersihkan dokumen dari komponen-komponen yang tidak memiliki hubungan dengan informasi yang ada pada dokumen, seperti, Karakter atau simbol (@#\$%^&\*()\_+”:{}<>.,?!~/[]), angka, link url (<http://link.com>), hashtag (#), dan mention @username.
2. Tahap *case folding*, yaitu untuk mengubah seluruh huruf yang ada pada dokumen menjadi huruf kecil.
3. Kalimat atau kata yang masuk hasil dari proses *case folding* akan dilakukan proses *tokenizing* yaitu memisahkan setiap kata di dalam dokumen yang awalnya merupakan sebuah kalimat menjadi dipisah perkata.
4. Selanjutnya proses *stemming* yaitu pengubahan kata menjadi kata dasar. Algoritma stemming yang digunakan adalah Algoritma ECS ( *Enhanced Confix Stripping*)
5. Masuk pada proses *normalisasi* yaitu mengembalikan kata-kata tidak baku menjadi baku sesuai dalam kamus Bahasa Indonesia.
6. Proses *Filtering* yaitu proses mengurangi jumlah kata-kata yang ada dengan penghilangan *stopword*. Dimana *stopword* adalah kata yang bukan merupakan ciri atau kata unik dari suatu dokumen seperti kata sambung dan kata kepunyaan.
7. Selanjutnya akan dilakukan proses *negation handling* yaitu proses penggabungan kata negasi dengan kata selanjutnya sehingga menjadi satu kata baru yang memiliki makna yang sama.
8. Proses *preprocessing* selesai.

### 3.4.3 Analisa Pembobotan (TF-IDF)

Pada tahapan selanjutnya setelah *preprocessing* yaitu dilakukan pembobotan kata. Data yang telah di *preprocessing* diubah ke bentuk *vector* atau *numerik* supaya mudah saat proses komputasi. Untuk mengubah data tersebut menjadi *numerik* bisa



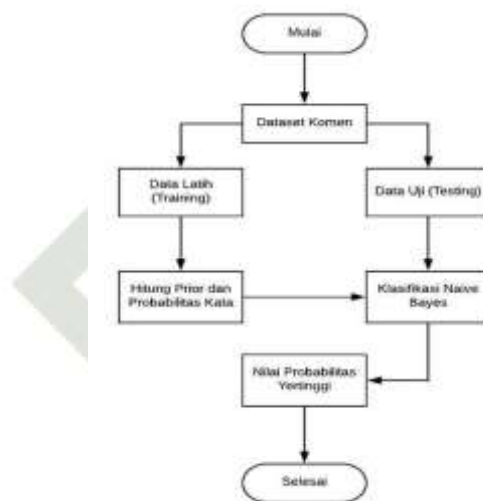
#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dengan proses pembobotan TF-IDF. Dimana pertama yaitu menghitung frekuensi kemunculan kata dalam dokumen (TF) dan banyaknya dokumen dimana suatu kata itu muncul (IDF), selanjutnya baru dilakukan penghitungan TF-IDF dengan menggunakan rumus.

#### 3.4.4 Analisa Klasifikasi Naive Bayes

Pada tahap ini Analisa Klasifikasi *Naive Bayes* dilakukan dari metode klasifikasi yang telah diperoleh dengan data latih/*Vocabulary* hasil observasi. Adapun langkah-langkah metode *Naive Bayes* yaitu:



**Gambar 3.2 Langkah-langkah Metode *Naive Bayes***

1. Bentuk *Vocabulary* data latih
2. Hitung nilai pembobotan masing-masing kata
3. Hitung nilai probabilitas untuk masing-masing kelas
4. Hitung nilai probabilitas untuk setiap kata terhadap masing-masing kelas
5. Hitung nilai probabilitas dokumen terhadap masing-masing kelas
6. Tentukan kelas dengan nilai probabilitas tertinggi

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3.5. Perancangan

Sampai ditahap ini aplikasi dirancang berdasarkan hasil analisa yang telah dilakukan pada proses sebelumnya. Tahapan dalam merancang sistem ini terdiri dari beberapa tahap yaitu.

#### 1. Perancangan Basis Data

Setelah sudah mendapatkan gambaran dari analisa pada aplikasi yang akan dibuat, maka selanjutnya dilakukan perancangan basis data untuk melengkapi komponen terhadap aplikasi yang akan dibuat.

#### 2. Perancangan Apikasi

Dalam perancangan aplikasi perlu adanya tampilan aplikasi yang baik. Karena perlu ditekankan merancang sebuah aplikasi haruslah memiliki tampilan yang baik dan mudah dimengerti dalam pengoperasiannya.

### 3.6. Implementasi dan Pengujian

Proses selanjutnya yaitu implementasi. Implementasi adalah proses yang dikembangkan berdasarkan hasil analisa dan perancangan yang telah dibuat. Pada tahap ini aplikasi telah selesai dan dapat dilakukan pengujian. Berikut merupakan penjelasan perangkat yang dibutuhkan dalam pembuatan aplikasi:

1. Bahasa Pemrograman : PHP
2. Database : MySQL
3. Browser : Google Chrome, Mozilla Firefox
4. Web Server : Apache

Setelah dilakukan implementasi maka dilakukanlah pengujian terhadap aplikasi. Pengujian dilakukan untuk mengetahui bagaimana kinerja metode. Dalam perhitungan tingkat keakurasian metodenya, akan digunakan pengujian *White Box* dan *Confussion Matriks*.

### 3.7. Kesimpulan dan Saran

Tahapan terakhir yaitu kesimpulan dan saran yang merupakan tahap penentuan terhadap hasil pengujian yang telah dilakukan. Dalam pengujian ini tujuannya adalah untuk menjawab rumusan masalah dan mengetahui bagaimana

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

hasil pengujian. Apakah penerapan algoritma *Naive Bayes* berhasil sehingga dapat bermanfaat. Selanjutnya, Pada tahap ini juga ditambahkan dengan saran yang berguna sebagai bahan pertimbangan dan gambaran untuk penelitian selanjutnya demi mendapatkan nilai akurasi yang lebih baik.





## BAB IV

### ANALISA DAN PERANCANGAN

Tahapan ini merupakan tahapan penting pada penelitian. Karena pada bab ini akan dilakukan uji analisa data, preproses data, pembobotan data dan proses klasifikasi dengan metode *Naive Bayes Classifier*.

#### 4.1. Analisa Data

Komentar-komentar di *twitter* pada Gubernur DKI Jakarta Aniesbaswedan tersebut nantinya akan dianalisa untuk diklasifikasikan apakah komentar tersebut termasuk kedalam komentar kategori positif atau negatif dengan menggunakan metode *Naive Bayes Classifier*. Komentar yang telah di kumpulkan akan diproses pada tahap *text preprocessing*. Berikut merupakan gambar alur tahap proses:



**Gambar 4.1 Umum Proses**

##### 4.1.1 Kebutuhan Data

Data yang dibutuhkan pada penelitian ini adalah, nama akun *twitter*, dan *teks tweet*. *tweet* yang digunakan sebagai data latih dan data uji dikumpulkan melalui trending topik yang membahas Gubernur DKI Jakarta Anies Baswedan. Data yang digunakan sebagai dataset yaitu total 600 komentar, dibagi menjadi 2

kelas yaitu positif, dan negatif, dengan 300 komentar di tiap kategorinya. Setelah itu komentar dilanjutkan ke proses pelabelan manual.

#### 4.1.2 Pengumpulan Data

Data input yang digunakan adalah *tweet* opini masyarakat terhadap Gubernur DKI Jakarta Anies Baswedan. Perbandingan *dataset* yang digunakan adalah 90% data latih : 10% data uji, 80% data latih : 20% data uji dan 70% data latih : 30% data uji. Tabel dibawah ini berisikan rincian *dataset* yang digunakan dan telah diberi kelas atau pelabelan manual:

**Tabel 4.1 Rincian Dataset 90% data latih : 10% data uji**

Data Komentar	Data Latih	540(90%)		
	Data Uji	60(10%)		
		Latih	Uji	Jumlah
Kelas	Positif	270	30	300
	Negatif	270	30	300
	Jumlah	540	60	600
		Twitt	Twitt	Twitt

**Tabel 4.2 Rincian Dataset 80% data latih : 20% data uji**

Data Komentar	Data Latih	540(80%)		
	Data Uji	60(20%)		
		Latih	Uji	Jumlah
Kelas	Positif	240	60	300
	Negatif	240	60	300
	Jumlah	480	120	600
		Twitt	Twitt	Twitt

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Tabel 4.3 Rincian Dataset 70% data latih : 30% data uji**

Data Komentar	Data Latih	540(70%)		
	Data Uji	60(30%)		
		Latih	Uji	Jumlah
Kelas	Positif	270	30	300
	Negatif	270	30	300
	Jumlah	540 Twitt	60 Twitt	600 Twitt

**Tabel 4.4 Contoh Data Komentar**

No	Twitter	Tweet	Kelas
1	@wisanggeni_084	Jgn banyak ngomong lah @anisbaswedan.....suruh noh ketua fetamburan....berani gak?	Negatif
2	@txtidaridreamzen @derbyeonz	Anis baswedan ngapain anjir ke inkigayo	Negatif
3	@eko_kuntadhi @arianarosaline	Jangan sampai doa saat maulid nabi di petamburan berumur pendek mantul ke anis baswedan. Berdoalah kau anis baswedan agar kau diberi umur yang panjang. Jangan kau percaya dengan ocehan gerombolan petamburan itu.	Negatif
4	@ndruwo_	Gak heran kalo anis baswedan positif covid 19 , jakarta kan paling top kasus covid nya . Pemerintah pusat harus bertindak tegas lagi untuk yang melanggar protokol Kesehatan	Positif
5	@lisaamartatara3	@aniesbaswedan cepat sembuh pak @anisbaswedan.. Semoga snantias dlm lindungan allah swt. Aamiin	Positif
6	@kafiradikalis	@aniesbaswedan lebih baik jujur walaupun itu menyakitkan.. Daripada	Positif



### Hak Cipta Dilindungi Undang-Undang

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		berbohong tapi tudak menyelesaikan masalah.. Get well soon pak @anisbaswedan	
7	@jpnncom	Dengan positifnya anis baswedan akan covid 19, ini membuktikan dia tidak disiplin. Boleh kau himbau agar masyarakat disiplin 3m, tapi semua orang mentertawakan karena anis itu orangnya "pintar omdo".	Negatif
8	@today.line	Anis kna covid,, bs jd c cabul jg covid,, mknx kabur.. @jayeng_seno_ @dtanjung15 @bhd_8 @agoesaguss @cagubnyinyir2 anies baswedan positif corona #linetoday #kumparan	Negatif
9	@ch_chotlmah @andiadi__	Lekas sembuh pak anis.. Semoga selalu dalam lindungan novel baswedan	Negatif
10	@m_asmara1701	Rezim ini yang dia bidik hanya habib rizieq shihab & anis baswedan karna takut kalau yang lain dibiarkan saja. Semoga rezim ini cepat dila'lanat.	Positif
11	@msaid_didu	@aniesbaswedan hebat mmg anisbaswedan sering mendapat penghargaan dari hasil karya orang lain	Negatif
12	@bemz_q @mamaze_asli	Andai handaru berdomisili di dki mungkin lain ceritanya...cc @anisbaswedan	Positif
13	@geiszchalifah	Saya share.... Karena aku suka tulisan dan kesan jujurnya ttg @anisbaswedan	Positif
14	@republikaonline	Pak bupati... Nggak usah gusar... Bapak aman.. Kan bukan di petamburan! Kan bapak bukan anis baswedan kan ulamanya bukan hrs kan bukan jakarta pokoake aman	Negatif
15	@marlina_idha	@aniesbaswedan @dkijakarta pada waktunya si pembenci akan terkubur dgn	Positif

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		prestasi.. Selamat buat pak @anisbaswedan dan @dkijakarta	
16	@santorinissun	Fadlizon, sebagai pejabat public layak dipidanakan krn mendatangi dan melanggar undang2 karantina di kediaman hrs. Contoh pejabat public terbodoh yg tdk memberikan contoh positive perilakunya saat pandemi. Sama seperti anisbaswedan juga layak dipidanakan @mohmahfudmd	Negatif
17	@fahiraidris	@aniesbaswedan bru d jaman anisbaswedan dpt penghargaan gubernur terpopuler gubernur2 yg dlu pda ngapain??? Anjing menggong khafilah tetap berlalu lanjutkan kekonyolan mu pakkk...biar mkin populer!!!	Positif
18	Rt @muham30205469: @b3doel_ @hamzah_saputraa	Anisbaswedan hanya tukang obat sama rizieq...	Negatif
19	@dennysiregar7	@kpk_ri @fadlizon saya sarankan dia @fadlizon agar berguru cara ngeles tingkat dewa sama @anisbaswedan	Negatif
20	@etty_markim	Anis baswedan sakit apa nyakit. ?. Anis baswedan satu - satunya manusia yg terpapar virus covid - 19 pada saat di isolasi masih bisa senyum - senyum dan cengengesan. Seperti tdk terjadi apa - apa.	Negatif
21	@ustadtengkuzul	Anda pejabat dan tengku boleh bersilat lidah sesukamu, memprofokasi sesukamu, ngomong sesukamu. Tapi klo yg diatas sudah berkehendak anda tidak bisa mengelak. Sekarang anis baswedan, trust reza patria, mungkin bentar lagi tengku ini nyusul gub & wagub dki... Biar tenant rakyat	Negatif
22	@ferdinandhaen3	Siapun yg melakukan tindak pidana korupsi harus diproses secara hukum kalau hanya isu saja, ini pembunuhan karakter sejumlah organisasi dan ormas	Positif

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

		masyarakat di jakarta siap membela gubernur anis baswedan demi keadilan dan kebenaran.	
23	@gabrielaazzare1	Dki prokese selain 3m sekarang menjadi 5m 1.memakai masker 2.mencuci tangan 3. Menjaga jarak 4. Menghindari kerumunan 5. Menjauhi anis baswedan.... Njir..meme djiancuk	Negatif
24	@bintang12898520	@mulyadimade @rachlannashidik @agusyudhoyono @sbyudhoyono kalau tdk dikerjain dari segala penjuru poling ahy paling tinggi waktu cagub dki,dan kasus baswedan yg disiram membuat anis jadi menang, yah begitulah politik dikita sekarang ,kwalitas ahy luar biasa dan sby tdk aji mumpung seoerti sekarang !jadi jangan assl jeplak !	Negatif
25	@geiszchalifah	Ini salah gubernur dki anis baswedan.. Kok medan sampai banjir.	Negatif
26	@yuqheixuxi	Oh itu ada yang jepotong ternyata chatnya. Ini loh yang aku respon jangan kayak anis baswedan ngomongnya wkwkwkwkwek.	Negatif
27	Rt @pejuangmuda152:	Silaturahmi memang berkah dan bagus, tapi harus tau kondisi juga. Dia jempuit sendiri virusnya, klaster baru virus coro...	Negatif
28	@kakekhalal	Semoga anis baswedan jadi presiden ri ke 2024..aminn aminnn	Positif
29	@geiszchalifah	Od itu artinya (orang dengki) lebih kuat ke situ sih sebenernya bukan (otak dikit), walaupun semaksimal apapun dan sebgus apa pun kinerja bpk.anis baswedan tetep di pandang jelek sama si (orang dengki).. Bukan nya begitu om? Hehehehe	Negatif
30	@elisa_jkt	Wah bisa tuh Pak @aniesbaswedan gratiskan vaksin utk non PBI di DKI ...	Negatif



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

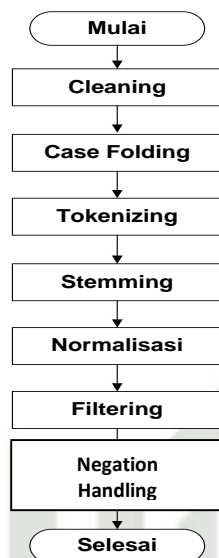
		Biar asik dikasih insentif 10rb, eh atau 33 ribu/orang ...	
31	Rt @dtanjung15:	Guys manusia2 ini ngakunya bela ham tapi faktanya bela anis baswedan dan kelompoknya karna ada dapat dana kalo benar manusi...	Negatif
32	@gaksopan	Anjing gua kira anis baswedan	Negatif
33	Rt @dtanjung15:	Guys, pasukan dan simpatisan ubur-ubur ngamuk2 sama nyai. Karna nyai mau kirimkan anis baswedan karangan bunga. Mereka berp...	Negatif
34	@geiszchalifah	@fadlizon maju terus jakarta ku. Semoga pak anis baswedan jd presiden secepat nya ammin..	Positif
35	@coeus_thetitan	Ketidak konsistensinya gubernur dki anis baswedan membuat penderita akibat terinfeksi virus covid19 di jakarta kembali meningkat.	Negatif
....	.....	.....	.....
...	.....	.....	.....
599	@marlina_idha	Kerenn, nih Sob Agro Edukasi Wisata Ragunan, JakSel.Salah satu Spot baru wisata pertanian di kota Jakarta	Positif
600	@Faqihabdulqodir	Inisih emejing banget, kota semotropolis jkt masih memperhatikan agro wisata.	Positif

#### 4.1.3 Text Preprocessing

Preproses (*Preprocessing*) atau tahapan preproses merupakan tahapan yang penting dalam melakukan analisa klasifikasi, bertujuan untuk membersihkan data-data dari unsur-unsur yang tidak dibutuhkan berguna untuk mempermudah dalam proses klasifikasi. Gambar 4.2 merupakan flowchart tahapan preproses data yang digunakan.

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



**Gambar 4.2 Flowchart Preproses Data**

Berikut adalah penjelasan dari Gambar 4.2 flowchart preproses data yang digunakan:

#### 1. *Cleaning*

*Cleaning* akan membersihkan atau menghilangkan karakter dan entitas pada komentar yang tidak diperlukan. Adapun kata atau karakter yang akan dihilangkan pada data komentar adalah karakter atau simbol (@#\$%^&\*<>![]), angka, *emoticon* dan *link* URL. Untuk lebih jelasnya dapat dilihat pada tabel 4.5 berikut:

**Tabel 4.5 Hasil *Cleaning* Data Komentar Gubernur DKI Jakarta**

No	Akun <i>tweet</i>	Komentar	Hasil <i>Cleaning</i>	Kelas
1	@KakekHalal	Semoga anis baswedan jadi presiden RI ke 2024..aminn aminnn.	Semoga jadi presiden RI ke aminn aminnn	Positif
2	@marlina_idha	Kabar baik untuk Warga Jakarta! Kebijakan pak Gub @aniesbaswedan Hapus Denda & Diskon Pajak DKI. Kendaraan 50% PBB 20% Enak kalo punya Gub Amanah	Kabar baik untuk Warga Jakarta Kebijakan pak Gub Hapus Denda Diskon Pajak DKI Kendaraan PBB Enak kalau punya Gub Amanah	Positif

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3	@ AlfredVP	Seragam nya perhatikan,,,,,masa DKI ga punya duit buat beli seragam petugasnya,,,,!!!!!!!!!!	Seragam nya perhatikan masa DKI ga punya duit buat beli seragam petugasnya	Negatif
---	------------	--	--	---------

## 2. Case Folding

Proses *case folding* adalah tahapan penyeragaman bentuk huruf dengan mengubah semua huruf yang ada menjadi huruf kecil. Tabel 4.6 adalah hasil data yang telah dilakukan proses *case folding*:

**Tabel 4.6 Hasil Case Folding Data Komentar Gubernur DKI Jakarta**

No	Akun tweet	Hasil Cleaning	Hasil Case Folding	Kelas
1	@KakekHalal	Semoga anis baswedan jadi presiden RI ke 2024..aminn aminnn.	semoga jadi presiden ri ke aminn aminnn	Positif
2	@marlina_idha	Kabar baik untuk Warga Jakarta! Kebijakan pak Gub @aniesbaswedan Hapus Denda & Diskon Pajak DKI. Kendaraan 50% PBB 20% Enak kalo punya Gub Amanah	kabar baik untuk warga jakarta kebijakan pak gub hapus denda diskon pajak dki kendaraan pbb enak kalau punya gub amanah	Positif
3	@ AlfredVP	Seragam nya perhatikan,,,,,masa DKI ga punya duit buat beli seragam petugasnya,,,,!!!!!!!!!!	seragam nya perhatikan masa dki ga punya duit buat beli seragam petugasnya	Negatif

## 3. Tokenizing

Proses memisahkan teks menjadi kata, dikenal sebagai *token*. Bertujuan untuk proses memisahkan atau memecah teks atau kalimat menjadi sebuah kumpulan kata agar mudah dalam melakukan pembobotan tiap kata nantinya. Hasil *tokenizing* untuk contoh data komentar dapat dilihat pada tabel 4.7:



**Tabel 4.7 Hasil *Tokenizing* Data Komentar Gubernur DKI Jakarta**

No	Akun tweet	Hasil <i>Case Folding</i>	Hasil <i>Tokenizing</i>	Kelas
1	@KakekHalal	Semoga anis baswedan jadi presiden RI ke 2024..aminnn aminnn.	semoga aminnn	Positif
			jadi	
			presiden	
			ri	
			ke	
			aminnn	
			aminnn	
2	@marlina_idha	Kabar baik untuk Warga Jakarta! Kebijakan pak Gub @aniesbaswedan Hapus Denda & Diskon Pajak DKI. Kendaraan 50% PBB 20% Enak kalo punya Gub Amanah	kabar	Positif
			baik	
			untuk	
			warga	
			jakarta	
			kebijaksanaan	
			pak	
			gub	
			hapus	
			denda	
			diskon	
			pajak	
			dki	
			kendaraan	
			pbb	
			enak	

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

© Hak Cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

			kalau	
			punya	
			gub	
			amanah	
3	@ AlfredVP	Seragam nya perhatikan,,,,,masa DKI ga punya duit buat beli seragam petugasnya,,,,,!!!!!!!	seragam	Negatif
			nya	
			hati	
			masa	
			dki	
			ga	
			punya	
			duit	
			buat	
			beli	
			seragam	
			petugasnya	

#### 4. Stemming

Langkah-langkah pada *stemming* menggunakan ECS ini yaitu dengan mengembalikan sebuah kata menjadi bentuk dasar dengan menghilangkan imbuhan, jika kata terdapat didalam kamus maka dikembalikan ke kata dasar. Namun, jika kata tidak ada didalam kamus, maka kata tidak dikembalikan. Berikut tabel 4.8 adalah hasil data yang dilakukan proses *stemming*.

**Tabel 4.8 Hasil *Stemming* Data Komentar Gubernur DKI Jakarta**

No	Akun tweet	Hasil <i>Tokenizing</i>	<i>Stemming</i>	Kelas
1	@KakekHalal	semoga	moga	Positif
		jadi	jadi	
		presiden	presiden	
		ri	ri	
		ke	ke	
		amin	amin	
		amin	amin	
2	@marlina_idha	kabar	kabar	Positif
		baik	baik	
		untuk	untuk	
		warga	warga	
		jakarta	jakarta	
		kebijaksanaan	bijaksana	
		pak	pak	
		gub	gub	
		hapus	hapus	
		denda	denda	
		diskon	diskon	
		pajak	pajak	
		dki	dki	
		kendaraan	kendara	
		pbb	pbb	
		enak	enak	

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

© Hak Cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3	@ AlfredVP	kalau	kalau	Negatif
		punya	punya	
		gub	gub	
		amanah	amanah	
3	@ AlfredVP	seragam	seragam	
		Nya	nya	
		Perhatikan	hati	
		Masa	masa	
		Dki	dki	
		Ga	ga	
		Punya	punya	
		Duit	duit	
		Buat	buat	
		Beli	beli	
		Seragam	seragam	
		petugasnya	tugas	

### 5. Normalisasi

Normalisasi merupakan proses yang bertujuan untuk mengubah kata yang tidak sesuai ejaan. Proses Untuk menangani seperti ini dilakukan dengan melibatkan kamus yang dibuat terdiri dari kata tak baku dan kata bakunya. Data pada penelitian ini yaitu twitt berbahasa Indonesia terdapat kata yang tak sesuai ejaan (kata tak baku). Terdapat beberapa ciri unik kata tak baku yang terdapat dalam twitt bahasa indonesia (Hidayatullah, 2014), ciri tersebut antara lain:

1. Menggunakan angka sebagai pembentuk kata seperti kata ‘masing2’ untuk mengungkapkan kata ‘masing-masing’

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Menghilangkan huruf untuk menyingkat kata, seperti kata tidak menjadi 'tdk' Berikut tabel 4.9 adalah hasil data yang dilakukan proses normalisasi.

Pada penelitian tersebut proses untuk menangani kata tak baku dilakukan dengan membuat sebuah kamus konversi kata yang terdiri dari kata tak baku dan kata bakunya. Proses dilakukan dengan melibatkan kamus tersebut untuk mengubah kata tak baku. Berikut contoh twitt yang melakukan proses normalisasi.

**Tabel 4.9 Hasil Normaliasi Data Komentar Gubernur DKI Jakarta**

No	Akun tweet	Hasil Stemming	Normalisasi	Kelas
1	@KakekHalal	moga	moga	Positif
		jadi	jadi	
		presiden	presiden	
		ri	republik	
			indonesia	
		ke	ke	
		aminn	amin	
2	@marlina_idha	aminn	amin	Positif
		kabar	kabar	
		baik	baik	
		untuk	untuk	
		warga	warga	
		jakarta	jakarta	
		bijaksana	bijaksana	
		pak	pak	
		gub	gubernur	
		hapus	hapus	
		denda	denda	

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3	@ AlfredVP	diskon	diskon	Negatif
		pajak	pajak	
		dki	dki	
		kendara	kendara	
		pbb	pbb	
		enak	enak	
		kalau	kalau	
		punya	punya	
		gub	gubernur	
		amanah	amanah	
3	@ AlfredVP	seragam	seragam	Negatif
		nya	nya	
		hati	hati	
		masa	masa	
		dki	dki	
		ga	tidak	
		punya	punya	
		duit	duit	
		buat	buat	
		beli	beli	
3	@ AlfredVP	seragam	seragam	Negatif
		petugasnya	tugas	

## 6. Filtering

Proses menghilangkan kata-kata yang dianggap tidak memiliki makna (*stopwords*). Alasan penghapusan *stopwords* dalam banyak program aplikasi yang



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

berkaitan dengan text mining adalah karena penggunaannya sangat umum sehingga pengguna dapat berfokus pada kata-kata yang jauh lebih penting (Ganesan, 2015).

Pembuatan kamus stopwords pada penelitian ini menggunakan kamus stopwords (stoplist) Tala yang telah dibentuk pada penelitian 'A Study Of Stemming Effects On Information Retrieval in Bahasa Indonesia' (Tala, 2003) dan menggunakan cara Tala mendapatkan stopwords. Pada penelitian tersebut, stoplist terdiri dari kata-kata yang paling umum digunakan dalam berita-berita beberapa koran berbahasa Indonesia. Tabel 4.10 adalah hasil data yang telah dilakukan proses *filtering*.

**Tabel 4.10 Hasil *Filtering* Data Komentar Gubernur DKI Jakarta**

No	Akun tweet	Hasil Normalisasi	<i>Filtering</i>	Kelas
1	@KakekHalal	moga	moga	Positif
		jadi	jadi	
		presiden	presiden	
		republik	republik	
		indonesia	indonesia	
		ke		
		amin	amin	
		amin		
2	@marlina_idha	kabar	kabar	Positif
		baik	baik	
		untuk		
		warga	warga	
		jakarta	jakarta	
		bijaksana	bijaksana	
		pak	bapak	

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3	@ AlfredVP	gubernur	gubernur	Negatif
		hapus		
		denda	denda	
		diskon	diskon	
		pajak	pajak	
		dki	dki	
		kendara	kendara	
		pbb	pbb	
		enak	enak	
		kalau		
		punya	punya	
		gubernur	gubernur	
3	@ AlfredVP	amanah	amanah	Negatif
		seragam	seragam	
		nya		
		hati	hati	
		masa	masa	
		dki	dki	
		tidak		
		punya	punya	
		duit	duit	
		buat		
		beli	beli	
		seragam	seragam	
		tugas	tugas	

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 7. *Negation Handling*

*Negation Handling* atau penanganan negasi merupakan mencari kata yang bersifat negasi, karena kata negasi dapat mengubah nilai positif atau negatif dari data, dengan cara mendeteksi kata negasi dan membalikkan kata setelah kata negasi tersebut dengan cara menggunakan kamus. Contohnya seperti kata “tidak baik” menjadi “buruk”. Berikut adalah hasil data yang telah dilakukan proses negation handling.

**Tabel 4.11 Hasil *Negation Handling* Data Komentar Gubernur DKI Jakarta**

No	Akun tweet	Hasil <i>Filtering</i>	<i>Negation Handling</i>	Kelas
1	@KakekHalal	moga	moga	Positif
		jadi	jadi	
		presiden	presiden	
		republik	republik	
		indonesia	indonesia	
		amin	amin	
2	@marlina_idha	kabar	kabar	Positif
		baik	baik	
		warga	warga	
		jakarta	jakarta	
		bijaksana	bijaksana	
		bapak	bapak	
		gubernur	gubernur	



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3	@ AlfredVP	denda	denda	Negatif
		diskon	diskon	
		pajak	pajak	
		dki	dki	
		kendara	kendara	
		pbb	pbb	
		enak	enak	
		punya	punya	
		gubernur	gubernur	
		amanah	amanah	
3	@ AlfredVP	seragam	seragam	Negatif
		hati	hati	
		masa	masa	
		dki	dki	
		punya	punya	
		duit	duit	
		beli	beli	
		seragam	seragam	
		tugas	tugas	

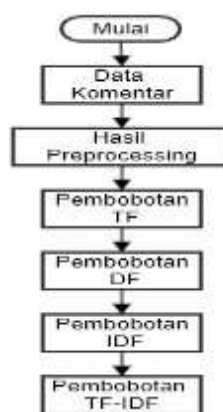
#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 4.1.4 Tahap Pembobotan (TF-IDF)

Pembobotan adalah proses mengubah kata menjadi bentuk angka atau vektor. Dan TF (*Term Frequency*) adalah frekuensi dari kemunculan term dalam dokumen yang bersangkutan. Semakin besar jumlah kemunculan suatu term dalam dokumen, maka akan semakin besar bobotnya atau akan memberikan nilai kesesuaian yang semakin besar.

Sedangkan TF-IDF merupakan metode untuk menghitung nilai *term frequency* (tf) dan *invers document frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Berikut proses pembobotan TF-IDF yang terdapat pada gambar 4.3 dibawah ini:



**Gambar 4.3 Flowchart pembobotan TF-IDF**

*Flowchart* diatas menjelaskan tahapan untuk mendapatkan nilai TF-IDF. Pada metode ini akan dihitung bobot setiap kata dalam dokumen dengan menggunakan rumus:

$$W_{dt} = tf_{dt} * IDF_t$$

Dimana:

d : dokumen ke-d

t : kata ke-t dari kata kunci

W : bobot dokumen ke-d terhadap kata ke-t

tf : banyak kata yang dicari pada sebuah dokumen

IDF : *inversed document frequency*

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Untuk nilai IDF itu sendiri didapatkan dari **IDF :  $\log_2(D/df)$**

Dimana:

D : total dokumen

df : banyak dokumen yang mengandung kata yang banyak dicari

1. Langkah pertama yang dilakukan menghitung TF dengan cara menghitung bobot tiap term dalam teks. Dengan memberikan bobot nilai 1 dan 0. **Tabel 4.12**

**Pembobotan TF**

Kata (Term)	TF		
	Data 1	Data 2	Data 3
moga	1	0	0
jadi	1	0	0
presiden	1	0	0
republik	1	0	0
indonesia	1	0	0
amin	2	0	0
kabar	0	1	0
baik	0	1	0
warga	0	1	0
suruh	0	1	0
jakarta	0	1	0
bijaksana	0	1	0
bapak	0	1	0
gubernur	0	2	0
denda	0	1	0
diskon	0	1	0
pajak	0	1	0
dki	0	1	0
kendara	0	1	0
pbb	0	1	0
enak	0	1	0
punya	0	1	0
amanah	0	1	0
seragam	0	0	1
hati	0	0	1
masa	0	0	1
punya	0	0	1
duit	0	0	1
beli	0	0	1
seragam	0	0	1
dinas	0	0	1

2. Langkah kedua menghitung DF dengan cara menjumlahkan bobot nilai term dalam dokumen.



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Tabel 4.30 Hasil Pembobotan DF**

Kata (Term)	TF			IDF
	Data 1	Data 2	Data 3	DF
moga	1	0	0	1
jadi	1	0	0	1
presiden	1	0	0	1
republik	1	0	0	1
indonesia	1	0	0	1
amin	2	0	0	1
kabar	0	1	0	1
baik	0	1	0	1
warga	0	1	0	1
suruh	0	1	0	1
jakarta	0	1	0	1
bijaksana	0	1	0	1
bapak	0	1	0	1
gubernur	0	2	0	1
denda	0	1	0	1
diskon	0	1	0	1
pajak	0	1	0	1
dki	0	1	0	1
kendara	0	1	0	1
pbb	0	1	0	1
enak	0	1	0	1
punya	0	1	0	1
amanah	0	1	0	1
seragam	0	0	1	1
hati	0	0	1	1
masa	0	0	1	1
punya	0	0	1	1
duit	0	0	1	1
beli	0	0	1	1
seragam	0	0	1	1
dinas	0	0	1	1

3. Langkah ketiga menghitung IDF dengan menggunakan rumus:

**IDF :  $\log_2(D/df)$ .**

Dimana:

D : total dokumen

df : banyak dokumen yang mengandung kata yang banyak dicari

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Tabel 4.14 Hasil Pembobotan IDF**

Kata (Term)	TF			IDF	
	Data 1	Data 2	Data 3	DF	log <sub>2</sub> (D/df)
moga	1	0	0	1	0.301029996
jadi	1	0	0	1	0.301029996
presiden	1	0	0	1	0.301029996
republik	1	0	0	1	0.301029996
indonesia	1	0	0	1	0.301029996
amin	2	0	0	1	0.301029996
kabar	0	1	0	1	0.301029996
baik	0	1	0	1	0.301029996
warga	0	1	0	1	0.301029996
suruh	0	1	0	1	0.301029996
jakarta	0	1	0	1	0.301029996
bijaksana	0	1	0	1	0.301029996
bapak	0	1	0	1	0.301029996
gubernur	0	2	0	1	0.301029996
denda	0	1	0	1	0.301029996
diskon	0	1	0	1	0.301029996
pajak	0	1	0	1	0.301029996
dki	0	1	0	1	0.301029996
kendara	0	1	0	1	0.301029996
pbb	0	1	0	1	0.301029996
enak	0	1	0	1	0.301029996
punya	0	1	0	1	0.301029996
amanah	0	1	0	1	0.301029996
seragam	0	0	1	1	0.301029996
hati	0	0	1	1	0.301029996
masa	0	0	1	1	0.301029996
punya	0	0	1	1	0.301029996
duit	0	0	1	1	0.301029996
beli	0	0	1	1	0.301029996
seragam	0	0	1	1	0.301029996
dinas	0	0	1	1	0.301029996

4. Langkah keempat menghitung bobot TF-IDF.

Pada tabel 4.15 adalah hasil pembobotan berdasarkan data komentar di tabel sebelumnya. Bobot akhir suatu term adalah mengalikan keduanya yaitu **TF x IDF**

**Tabel 4.15 Hasil Pembobotan TF-IDF**

Kata (Term)	TF			IDF				
	Data 1	Data 2	Data 3	DF	$\log_2(D/df)$	Data 1	Data 2	Data 3
moga	1	0	0	1	0.301029996	0.301029996	0	0
jadi	1	0	0	1	0.301029996	0.301029996	0	0
presiden	1	0	0	1	0.301029996	0.301029996	0	0
republik	1	0	0	1	0.301029996	0.301029996	0	0
indonesia	1	0	0	1	0.301029996	0.301029996	0	0
amin	2	0	0	1	0.301029996	0.602059991	0	0
kabar	0	1	0	1	0.301029996	0	0.30103	0
baik	0	1	0	1	0.301029996	0	0.30103	0
warga	0	1	0	1	0.301029996	0	0.30103	0
suruh	0	1	0	1	0.301029996	0	0.30103	0
jakarta	0	1	0	1	0.301029996	0	0.30103	0
bijaksana	0	1	0	1	0.301029996	0	0.30103	0
bapak	0	1	0	1	0.301029996	0	0.30103	0
gubernur	0	2	0	1	0.301029996	0	0.30103	0
denda	0	1	0	1	0.301029996	0	0.30103	0
diskon	0	1	0	1	0.301029996	0	0.30103	0
pajak	0	1	0	1	0.301029996	0	0.30103	0
dki	0	1	0	1	0.301029996	0	0.30103	0
kendara	0	1	0	1	0.301029996	0	0.30103	0
pbb	0	1	0	1	0.301029996	0	0.30103	0
enak	0	1	0	1	0.301029996	0	0.30103	0
punya	0	1	0	1	0.301029996	0	0.30103	0
amanah	0	1	0	1	0.301029996	0	0.30103	0
seragam	0	0	1	1	0.301029996	0	0	0.30103
hati	0	0	1	1	0.301029996	0	0	0.30103
masa	0	0	1	1	0.301029996	0	0	0.30103
punya	0	0	1	1	0.301029996	0	0	0.30103
duit	0	0	1	1	0.301029996	0	0	0.30103
beli	0	0	1	1	0.301029996	0	0	0.30103
seragam	0	0	1	1	0.301029996	0	0	0.30103
dinas	0	0	1	1	0.301029996	0	0	0.30103

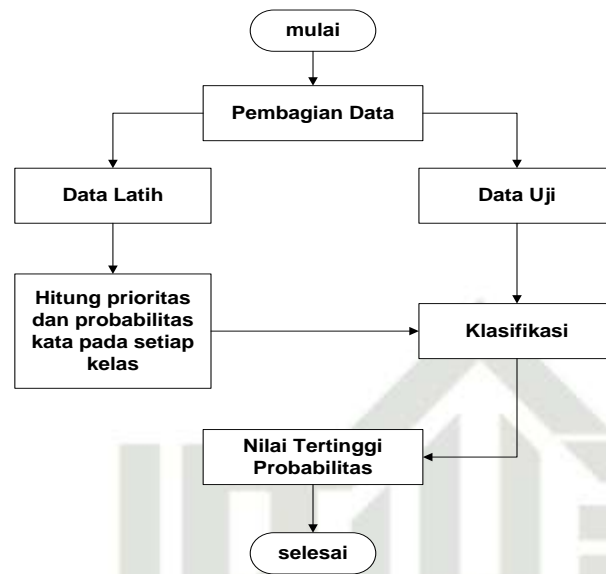
#### 4.1.5 Proses Klasifikasi Metode Naive Bayes Classifier

Pada tahapan ini akan dijelaskan teknik-teknik yang akan digunakan dalam pengklasifikasian data. *Flowchart* proses *Naive Bayes Classifier* dalam klasifikasi penelitian ini akan diperlihatkan pada gambar 4.4.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



**Gambar 4.4 Flowchart Proses Klasifikasi**

Pada proses klasifikasi Naive Bayes ini, seluruh dataset dibagi menjadi beberapa tahap data latih dan data uji. Berikut adalah penjelasan langkah-langkahnya beserta contoh perhitungan:

#### 1. Tahap data latih

Pada tahap ini, setiap kata yang telah diketahui nilai bobot TF-IDF nya akan dijadikan sebagai acuan untuk mencari nilai probabilitas kategori dan probabilitas masing-masing kata pada setiap *term* untuk setiap kelas dari data latih. Berikut adalah perhitungannya:

- a. Pertama hitung probabilitas setiap kategori (*prior*). pada penelitian ini yang menjadi kategori ada dua yaitu kategori positif, dan negatif.

$$P(\text{positif/negatif}) = \frac{D(\text{positif/negatif})}{|C|}$$

$$P(\text{Pos}) = \frac{d(\text{Pos})}{|c|} = 1/2$$

$$P(\text{Neg}) = \frac{d(\text{Neg})}{|c|} = 1/2$$

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Keterangan :

$p(pos/neg)$  = Peluang kemunculan dokumen dari kelas positif, dan negatif  
 $d(pos/neg)$  = Jumlah seluruh dokumen pada setiap kelas  
 $|c|$  = Jumlah keseluruhan data yang dilatih yaitu positif dan negatif

- b. Setelah melakukan perhitungan probabilitas setiap kategori kemudian selanjutnya menghitung probabilitas setiap *term* dari semua dokumen. Banyak *term* tergantung pada hasil preproses data. Berikut adalah merupakan perhitungan probabilitas pada setiap *term*:

$$P(w|positif/negatif) = \frac{(nk|positif/negatif) + 1}{(n, |pos/neg) + |kosakata|}$$

Keterangan :

$p(w_k|pos/neg)$  = Peluang kemunculan kata pada kategori  
 $w_k$  = Kata yang muncul pada sebuah kategori.  
 $(n_k, pos/neg) + 1$  = Jumlah frekuensi kemunculan kata pada kategori.  
 $n_k$  = Kemunculan setiap kata pada kategori.  
 $|kosakata|$  = Jumlah semua kata dari semua kategori.

- c. Probabilitas kata “moga”

$$p("moga"|"pos") = \frac{("moga"|"pos") + 1}{("pos") + |kosakata|} = \frac{0,301029996 + 1}{6 + 30} = 0,068475262929683$$

$$p("moga"|"neg") = \frac{("moga"|"neg") + 1}{("neg") + |kosakata|} = \frac{0 + 1}{7 + 30} = 0,05$$

- d. Probabilitas kata “jadi”

$$p("jadi"|"pos") = \frac{("jadi"|"pos") + 1}{("pos") + |kosakata|} = \frac{0,301029996 + 1}{6 + 30} = 0,068475262929683$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("jadi"|"neg") = \frac{("jadi"|"neg") + 1}{("neg") + |kosakata|} = \frac{0 + 1}{7 + 30} = 0,05$$

e. Probabilitas kata “presiden”

$$p("presiden"|"pos") = \frac{("presiden"|"pos") + 1}{("pos") + |kosakata|} = \frac{0,301029996 + 1}{6 + 30} = 0,068475262929683$$

$$p("presiden"|"neg") = \frac{("presiden"|"neg") + 1}{("neg") + |kosakata|} = \frac{0 + 1}{7 + 30} = 0,05$$

f. Probabilitas kata “republik”

$$p("republik"|"pos") = \frac{("republik"|"pos") + 1}{("pos") + |kosakata|} = \frac{0,301029996 + 1}{6 + 30} = 0,068475262929683$$

$$p("republik"|"neg") = \frac{("republik"|"neg") + 1}{("neg") + |kosakata|} = \frac{0 + 1}{7 + 30} = 0,05$$

g. Probabilitas kata “indonesia”

$$p("indonesia"|"pos") = \frac{("indonesia"|"pos") + 1}{("pos") + |kosakata|} = \frac{0,301029996 + 1}{6 + 30} = 0,068475262929683$$

$$p("indonesia"|"neg") = \frac{("indonesia"|"neg") + 1}{("neg") + |kosakata|} = \frac{0 + 1}{7 + 30} = 0,05$$

h. Probabilitas kata “amin”

$$p("amin"|"pos") = \frac{("amin"|"pos") + 1}{("pos") + |kosakata|} = \frac{0,60206 + 1}{6 + 30} = 0,084318947$$

$$p("amin"|"neg") = \frac{("amin"|"neg") + 1}{("neg") + |kosakata|} = \frac{0 + 1}{7 + 30} = 0,05$$

i. Probabilitas kata “kabar”

$$p("jangan"|"pos") = \frac{("jangan"|"pos") + 1}{("pos") + |kosakata|} = \frac{0 + 1}{6 + 30} = 0,052631576$$



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p("jangan"|"neg") = \frac{("jangan"|"neg") + 1}{("neg") + |kosakata|} = \frac{0,301029996 + 1}{7 + 30} = 0.0650515$$

j. Probabilitas kata “banyak”

$$p("banyak"|"pos") = \frac{("banyak"|"pos") + 1}{("pos") + |kosakata|} = \frac{0 + 1}{6 + 30} = 0,052631576$$

$$p("banyak"|"neg") = \frac{("banyak"|"neg") + 1}{("neg") + |kosakata|} = \frac{0,301029996 + 1}{7 + 30} = 0.0650515$$

k. Probabilitas kata “omong”

$$p("omong"|"pos") = \frac{("omong"|"pos") + 1}{("pos") + |kosakata|} = \frac{0 + 1}{6 + 30} = 0,052631576$$

$$p("omong"|"neg") = \frac{("omong"|"neg") + 1}{("neg") + |kosakata|} = \frac{0,301029996 + 1}{7 + 30} = 0.0650515$$

l. Probabilitas kata “suruh”

$$p("suruh"|"pos") = \frac{("suruh"|"pos") + 1}{("pos") + |kosakata|} = \frac{0 + 1}{6 + 30} = 0,052631576$$

$$p("suruh"|"neg") = \frac{("suruh"|"neg") + 1}{("neg") + |kosakata|} = \frac{0,301029996 + 1}{7 + 30} = 0.0650515$$

m. Probabilitas kata “ketua”

$$p("ketua"|"pos") = \frac{("ketua"|"pos") + 1}{("pos") + |kosakata|} = \frac{0 + 1}{6 + 30} = 0,052631576$$

$$p("ketua"|"neg") = \frac{("ketua"|"neg") + 1}{("neg") + |kosakata|} = \frac{0,301029996 + 1}{7 + 30} = 0.0650515$$

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

n. Probabilitas kata “berani”

$$p(\text{"berani"}|\text{"pos"}) = \frac{(\text{"berani"}|\text{"pos"}) + 1}{(\text{"pos"}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 30} = \mathbf{0,052631576}$$

$$p(\text{"berani"}|\text{"neg"}) = \frac{(\text{"berani"}|\text{"neg"}) + 1}{(\text{"neg"}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{7 + 30} \\ = \mathbf{0.0650515}$$

o. Probabilitas kata “tidak”

$$p(\text{"tidak"}|\text{"pos"}) = \frac{(\text{"tidak"}|\text{"pos"}) + 1}{(\text{"pos"}) + |\text{kosakata}|} = \frac{0 + 1}{6 + 30} = \mathbf{0,052631576}$$

$$p(\text{"tidak"}|\text{"neg"}) = \frac{(\text{"tidak"}|\text{"neg"}) + 1}{(\text{"neg"}) + |\text{kosakata}|} = \frac{0,301029996 + 1}{7 + 30} \\ = \mathbf{0.0650515}$$

## 2. Tahap Testing

Tahap ini merupakan tahapan setelah proses *training* yang telah dilakukan diatas. Pada tahapan ini dilakukan pengujian dengan cara menggunakan data uji kedalam model yang telah dibentuk pada tahapan *training* diatas.

**Tabel 4.14 Data testing**

No	Twitt	Kelas
1	Penghargaan ini semoga jadi penyemangat bapak saat ini...	Positif
2	Jgn banyak ngomong lah @anisbaswedan.....suruh noh Ketua fetamburan. berani gak?"	Negatif

Proses selanjutnya yang akan dilakukan adalah menghitung nilai probabilitas berdasarkan nilai setiap *term*. Perhitungan nilai probabilitas dilakukan dengan mengalikan nilai probabilitas semua kategori dengan probabilitas tiap *term* yang diambil dari semua data. Sebelum melakukan perhitungan (*testing*) sebelumnya data *testing* harus melalui tahapan *preprocessing* terlebih dahulu. Berikut hasil tahapan *preprocessing* dari data *testing*.

**Tabel 4.15 Data setelah proses *Preprocessing***

No	Twitt	Kelas
1	harga moga jadi semangat bapak saat sakit amin	Positif
2	jangan banyak omong suruh ketua berani tidak	Negatif

Tahapan selanjutnya, akan melakukan probabilitas masing-masing kata didata uji dengan menggunakan hasil probabilitas *term* yang didapat pada data latih.

**Tabel 4.16 Data latih dan data uji yang telah melalui tahapan *Preprocessing***

Data latih	Twitt	Kelas
1	moga jadi presiden republik indonesia amin amin	Positif
2	jangan banyak omong suruh ketua berani tidak	Negatif
Data uji		
3	moga jadi semangat bapak saat sakit amin jangan banyak omong suruh ketua berani tidak	?

Selanjutnya dilakukan perhitungan probabilitas masing-masing kata diproses dengan menggunakan persamaan 2.3

$$p(w_k|pos/neg) = \frac{(n_k, pos/neg) + 1}{(n, pos/neg) + |kosakata|}$$

**Tabel 4.1 Pembobotan data uji**

No	Kata	TF 1	TF 2	DF	IDF	TF-IDF	TF-IDF
1	moga	1	0	1	0.301029996	0.301029996	0
2	jadi	1	0	1	0.301029996	0.301029996	0
3	amin	1	0	1	0.301029996	0.301029996	0
4	jangan	0	1	1	0.301029996	0	0.301029996
5	banyak	0	1	1	0.301029996	0	0.301029996



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

6	omong	0	1	1	0.301029996	0	0.301029996
7	suruh	0	1	1	0.301029996	0	0.301029996
8	ketua	0	1	1	0.301029996	0	0.301029996
9	berani	0	1	1	0.301029996	0	0.301029996
10	Tidak	0	1	1	0.301029996	0	0.301029996

Diketahui :

Total nilai TF [kosakata] = 10

Total nilai TF Positif = 3

Total nilai TF Negatif = 7

#### Probabilitas Data uji (Positif)

$$p(\text{"moga"}|\text{"pos"}) = (0.301029996 + 1)/(7 + 30) = 0,083014375$$

$$p(\text{"jadi"}|\text{"pos"}) = (0.301029996 + 1)/(7 + 30) = 0,083014375$$

$$p(\text{"amin"}|\text{"pos"}) = (0.301029996 + 1)/(7 + 30) = 0,083014375$$

$$p(\text{"jangan"}|\text{"pos"}) = (0 + 1)/(7 + 30) = 0.0625$$

$$p(\text{"banyak"}|\text{"pos"}) = (0 + 1)/(7 + 30) = 0.0625$$

$$p(\text{omong}|\text{"pos"}) = (0 + 1)/(7 + 30) = 0.0625$$

$$p(\text{"suruh"}|\text{"pos"}) = (0 + 1)/(7 + 30) = 0.0625$$

$$p(\text{"ketua"}|\text{"pos"}) = (0 + 1)/(7 + 30) = 0.0625$$

$$p(\text{"berani"}|\text{"pos"}) = (0 + 1)/(7 + 30) = 0.0625$$

$$p(\text{"tidak"}|\text{"pos"}) = (0 + 1)/(7 + 30) = 0.0625$$

#### Probabilitas Data uji (Negatif)

$$p(\text{"moga"}|\text{"neg"}) = (0 + 1)/(7 + 30) = 0,05$$

$$p(\text{"jadi"}|\text{"neg"}) = (0 + 1)/(7 + 30) = 0,05$$

$$p(\text{"amin"}|\text{"neg"}) = (0 + 1)/(7 + 30) = 0,05$$

$$p(\text{"jangan"}|\text{"neg"}) = (0.301029996 + 1)/(7 + 30) = 0.0650515$$

$$p(\text{"banyak"}|\text{"neg"}) = (0.301029996 + 1)/(7 + 30) = 0.0650515$$

$$p(\text{omong}|\text{"neg"}) = (0.301029996 + 1)/(7 + 30) = 0.0650515$$

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$p(\text{"suruh"}|\text{"neg"}) = (0.301029996 + 1)/(7 + 30) = 0.0650515$$

$$p(\text{"ketua"}|\text{"neg"}) = (0.301029996 + 1)/(7 + 30) = 0.0650515$$

$$p(\text{"berani"}|\text{"neg"}) = (0.4777121255 + 1)/(7 + 30) = 0.0650515$$

$$p(\text{"tidak"}|\text{"neg"}) = (0.4777121255 + 1)/(7 + 30) = 0.0650515$$

Tahapan selanjutnya menghitung probabilitas kategori

$$p(\text{pos/neg}|d) = p(\text{pos/neg}) * \prod_i P(a_i|\text{pos/neg})$$

Keterangan :

$p(\text{pos/neg})$  = Peluang kemunculan kata pada kategori atau kelas dengan dokumen data uji.

$d$  = Dokumen data baru

$\prod_i P(a_i|\text{pos/neg})$  = Peluang kemunculan  $a_i$  pada kategori atau kelas

$a_i$  = Kata baru yang akan diuji.

#### Probabilitas kategori positif

$$P(\text{uji}|\text{pos}) = p(\text{pos}) \times p(\text{moga}|\text{pos}) \times p(\text{jadi}|\text{pos}) \times p(\text{amin}|\text{pos})$$

$$p(\text{uji}|\text{pos}) = 0,5 \times 0,083014375 \times 0,083014375 \times 0,083014375$$

$$= \mathbf{0,000268826442454134}$$

$$P(\text{uji}|\text{pos}) = p(\text{pos}) \times p(\text{jangan}|\text{pos}) \times p(\text{banyak}|\text{pos}) \times p(\text{omong}|\text{pos}) \times p(\text{suruh}|\text{pos}) \\ \times p(\text{ketua}|\text{pos}) \times p(\text{berani}|\text{pos}) \times p(\text{tidak}|\text{pos})$$

$$p(\text{uji}|\text{pos})$$

$$= 0,5 \times 0,0625 \times 0,0625 \times 0,0625 \times 0,0625 \times 0,0625 \times 0,0625 \times 0,0625$$

$$= \mathbf{0,00000000186265}$$

#### Probabilitas kategori negatif

$$P(\text{uji}|\text{neg}) = p(\text{neg}) \times p(\text{moga}|\text{neg}) \times p(\text{jadi}|\text{neg}) \times p(\text{amin}|\text{neg})$$

$$p(\text{uji}|\text{neg}) = 0,5 \times 0,05 \times 0,05 \times 0,05$$

$$= \mathbf{0,0000625}$$

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$\begin{aligned}
 P(uji|neg) &= p(neg) \times p(jangan|neg) \times p(banyak|neg) \times p(omong|neg) \times p(suruh|neg) \\
 &\times p(ketua|neg) \times p(berani|neg) \times p(tidak|neg) \\
 p(uji|neg) &= 0,5 \times 0,0650515 \times 0,0650515 \times 0,0650515 \times 0,0650515 \times 0,0650515 \times 0,0650515 \\
 &\times 0,0650515 = \mathbf{0.00000000246474}
 \end{aligned}$$

Nilai probabilitas tertinggi adalah pada kategori negatif dengan nilai sebesar **0,00000000246474**, sehingga dapat diklasifikasikan kedalam kelas “**Negatif**”.

Nilai probabilitas tertinggi yaitu sebesar **0,000268826442454134** pada  $p(uji|pos)$ , sehingga komentar tersebut diklasifikasikan kedalam kelas “**Positif**”.

## 4.2. Perancangan

Perancangan klasifikasi positif, dan negatif berdasarkan komentar yang ada di *twitter* dirancang berdasarkan tahapan analisa yang dilakukan sebelumnya. Adapun bagian perancangan dilakukan sebagai acuan dalam pembuatan aplikasi dipenelitian ini, terdiri dari perancangan basis data (*database*), perancangan stuktur menu dan perancangan antar muka (*interface*).

### 4.2.1 Perancangan Basis Data (*Database*)

Perancangan basis data sistem terdiri dari 8 tabel yaitu *failed\_jobs*, *migrations*, *password\_resets*, *pembagian\_datas*, *pembagian\_datas\_detail*, *stopwords*, *tweets*, dan *users*.

**Tabel 4.15 Nama Tabel Database Gubernur DKI Jakarta**

No	Nama Tabel
1	<i>failed_jobs</i>
2	<i>migrations</i>
3	<i>password_resets</i>
4	<i>pembagian_datas</i>



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5	pembagian_datas_detail
6	stopwords
7	tweets
8	users

Berikut adalah struktur pada tabel *failed\_jobs* yang terdiri dari 7 *field*, seperti pada tabel 4.16:

**Tabel 4.16 Struktur Tabel *failed\_jobs***

No	Nama Field	Type dan Length	Primary Key
1	id	bigint(20)	YES
2	uuid	varchar (255)	YES
3	connection	text	
4	queue	text	
5	payload	longtext	
6	exception	longtext	
7	failed_at	timestamp	

Berikut adalah struktur pada tabel *migration* yang terdiri dari 3 tabel 4.17:

**Tabel 4.17 Struktur *Migration***

No	Nama Field	Type dan Length	Primary Key
1	id	bigint (20)	YES
2	Migration	Varchar(255)	
3	Batch	Int(11)	

Berikut adalah struktur pada tabel *password\_resets* yang terdiri dari 3 *field*, seperti yang terlihat pada tabel 4.18:

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Tabel 4.18 Struktur Tabel *Password\_reset***

No	Nama Field	Type dan Length	Primary Key
1	email	Varchar(255)	YES
2	token	varchar (255)	-
3	Created_at	timestamp	-

Berikut adalah struktur pada tabel *Stopwords*, yang terdiri dari 4 *field*, seperti yang terlihat pada tabel 4.19:

**Tabel 4.19 Struktur Tabel *Stopwords***

No	Nama Field	Type dan Length	Primary Key
1	id	Int(10)	YES
2	Text	Varchar (30)	-
3	Create_at	Timestamp	-
4	Update_at	Timestamp	-

Berikut adalah struktur pada tabel pembagian data yang terdiri dari 9 *field*, seperti yang terlihat pada tabel 4.20:

**Tabel 4.20 Struktur Tabel *Pembagian\_datas***

No	Nama Field	Type dan Length	Primary Key
1	id	Bigint(20)	YES
2	latih	Varchar(255)	-
3	Uji	Varchar(255)	-
4	Create_at	Timestamp	-
5	Update_at	Timestamp	-
6	Latih_positif	Int(11)	-
7	Uji_positif	Int(11)	-
8	Latih_negatif	Int(11)	-
9	Uji_negatif	int(11)	-

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Berikut adalah struktur pada tabel pembagian data detail yang terdiri dari 9 *field*, seperti yang terlihat pada tabel 4.22:

**Tabel 4.22 Struktur Tabel Pembagian\_data\_detail**

No	Nama Field	Type dan Length	Primary Key
1	id	int(11)	YES
2	akun	varchar(255)	-
3	tweet	text	-
4	kelas	Varchar(255)	-
5	Klasifikasi	Varchar(255)	
6	Created_at	timestamp	
7	Update_at	timestamp	
8	Id_pembagian_data	Int(11)	
9	Status	int(11)	

Berikut adalah struktur pada tabel *users* yang terdiri dari 8 *field*, seperti yang terlihat pada tabel 4.23:

**Tabel 4.23 Struktur Tabel Users**

No	Nama Field	Type dan Length	Primary Key
1	id	bigint(20)	YES
2	Name	Varchar (255)	-
3	Email	Varchar(255)	YES
4	Email_verified_at	timestamp	
5	Password	Varchar(255)	
6	Remember_token	Varchar(100)	



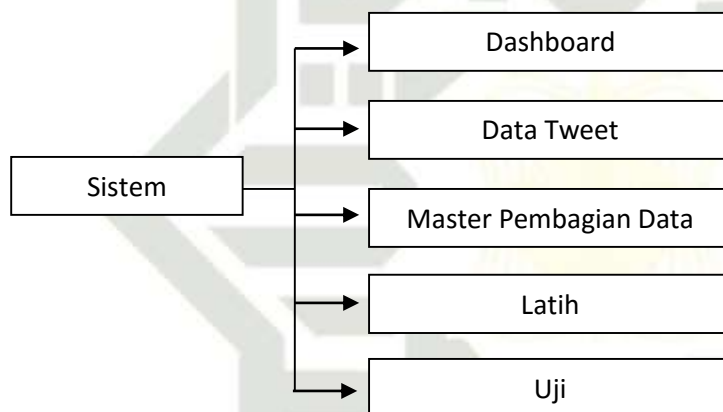
#### Hak Cipta Dilindungi Undang-Undang

1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Diarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

7	Created_at	timestamp	
8	Upload_at	timestamp	

#### 4.2.2 Perancangan Struktur Menu

Perancangan struktur menu menjelaskan hubungan halaman satu dengan halaman yang lainnya. Struktur menu adalah salah satu bagian penting untuk merancang sebuah sistem antar muka, karena akan mudah melihat tampilan dari sistem yang dibangun. Rancangan struktur menu dapat dilihat pada gambar 4.4 berikut:



**Gambar 4.4 Rancangan Struktur Menu**

#### 4.2.3 Perancangan Antarmuka (*Interface*)

Perancangan Antarmuka aplikasi bisa disebut sebagai sarana pengembangan yang digunakan untuk membuat komunikasi yang baik antara aplikasi dengan penggunanya. Di perancangan antarmuka ini dapat menerima informasi dari pengguna (*user*) dan memberikan informasi kepada pengguna (*user*) untuk membantu mengarahkan cara kerja penelusuran masalah sampai ditemukannya solusi terhadap masalah tersebut. Berikut ini merupakan tampilan antar muka yang akan dirancang pada penelitian ini.

## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 1. Rancangan Antar Muka Pada Halaman *Dashboard*

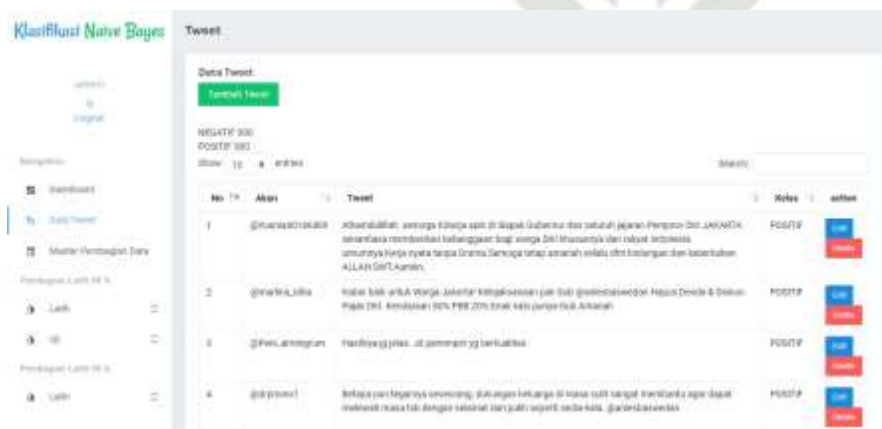
Halaman *Dashboard* merupakan halaman utama yang akan diakses dan ditampilkan aplikasi kepada pengguna. Halaman ini berisi tentang informasi judul penelitian, nama dan nim. Berikut ini adalah gambar 4.5 rancangan antarmuka halaman *Dashboard*:



Gambar 4.5 Rancangan Antarmuka Halaman *Dashboard*

### 2. Rancangan Antarmuka Pada Halaman Data *Tweet*

Halaman Data *Tweet* merupakan halaman yang berisikan informasi tentang komentar atau opini masyarakat tentang gubernur DKI Jakarta anise baswedan. Pada halaman ini juga menampilkan nama akun, data *tweet*, kelas serta terdapat tombol untuk menambah data *tweet*, edit data *tweet* dan hapus data *tweet*.



Gambar 4.6 Rancangan Antarmuka Halaman Data *Tweet*

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3. Rancangan Antarmuka Pada Halaman Master Pembagian Data *Tweet*

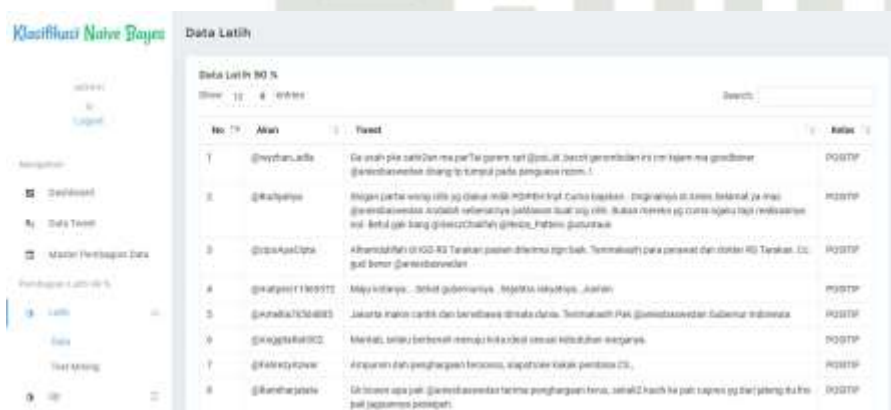
Halaman ini menampilkan data yang sudah di preproses tadi untuk selanjutnya dilakukan proses data latih menggunakan metode *Naive Bayes* yang proses data latihnya terdiri dari 3 yaitu untuk proses data latih 70%, 80% dan 90% ataupun bisa di sesuaikan oleh pengguna.



Gambar 4.7 Rancangan Antarmuka Halaman Master Pembagian Data

### 4. Rancangan Antarmuka Pada Halaman Data Latih

Pada Halaman ini berisi data latih yang nanti akan di jadikan sebagai bahan untuk data uji.



Gambar 4.8 Rancangan Antarmuka Halaman Data Latih

### 5. Rancangan Antarmuka Pada Halaman Data Uji



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Halaman ini sistem akan melakukan uji pada data apakah sesuai hasilnya dengan yang ada di data latih.

**Klasifikasi Naïve Bayes**

**Data Uji**

Data Latih 2:

No	Alasan	Tanggapan	Kelas
1	@kurnia444444	Alhamdulillah, semoga koreksi dari Bapak Ridwan dan seluruh jajaran Pimpinan DPT Jember bisa memberikan koreksi yang baik untuk kami. Kami akan berusaha untuk memperbaiki diri kami.	POSITIF
2	@kurnia_444	Kami akan berusaha untuk memperbaiki diri kami. Kami akan berusaha untuk memperbaiki diri kami.	POSITIF
3	@kurnia_444	Kami akan berusaha untuk memperbaiki diri kami. Kami akan berusaha untuk memperbaiki diri kami.	POSITIF
4	@kurnia_444	Kami akan berusaha untuk memperbaiki diri kami. Kami akan berusaha untuk memperbaiki diri kami.	POSITIF
5	@kurnia_444	Kami akan berusaha untuk memperbaiki diri kami. Kami akan berusaha untuk memperbaiki diri kami.	POSITIF
6	@kurnia_444	Kami akan berusaha untuk memperbaiki diri kami. Kami akan berusaha untuk memperbaiki diri kami.	POSITIF
7	@kurnia_444	Kami akan berusaha untuk memperbaiki diri kami. Kami akan berusaha untuk memperbaiki diri kami.	POSITIF

Gambar 4.16 Rancangan Antarmuka Halaman Data Uji

## BAB V

### IMPLEMENTASI DAN PENGUJIAN

#### 5.1. Implementasi

Implementasi adalah tahapan yang dilakukan untuk menerapkan perancangan yang telah dilakukan pada tahapan sebelumnya. Pada tahap implementasi ini akan dilihat apakah aplikasi telah berjalan sesuai dengan fungsinya atau tidak. Berikut adalah pembahasan yang akan dijelaskan pada tahapan implementasi:

##### 5.1.1 Lingkungan Implementasi

Dasarnya setiap desain aplikasi yang dirancang memerlukan sarana pendukung untuk memudahkan dalam menyelesaikan aplikasi yang akan dibuat. Sarana yang dimaksud seperti *hardware* yaitu kebutuhan perangkat keras komputer dan *software* yang digunakan untuk mengoperasikan aplikasi yang dibuat. Berikut ini beberapa sarana pendukung yang digunakan untuk mengimplementasikan aplikasi yang akan didesain yaitu:

##### 1. Perangkat Keras

*Prosesor* : Intel Core i5-7200U

*Memory* : 4 GB DDR4

##### 2. Perangkat Lunak

*Sistem operasi* : Windows 10

*Bahasa Pemrograman* : PHP

*Tools* : Notepad ++, Microsoft Visio

*Web Server* : Apache

*Web Browser* : Mozilla Firefox

*Database* : MySQL

##### 5.1.2 Implementasi Aplikasi

Tahapan implementasi Aplikasi ini merupakan tahap implementasi hasil perancangan antarmuka yang telah dibuat. Berikut ini merupakan implementasi aplikasi klasifikasi sentimen berdasarkan komentar *Anies baswedan*.

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 1. Implementasi Antarmuka Halaman *Dashboard*

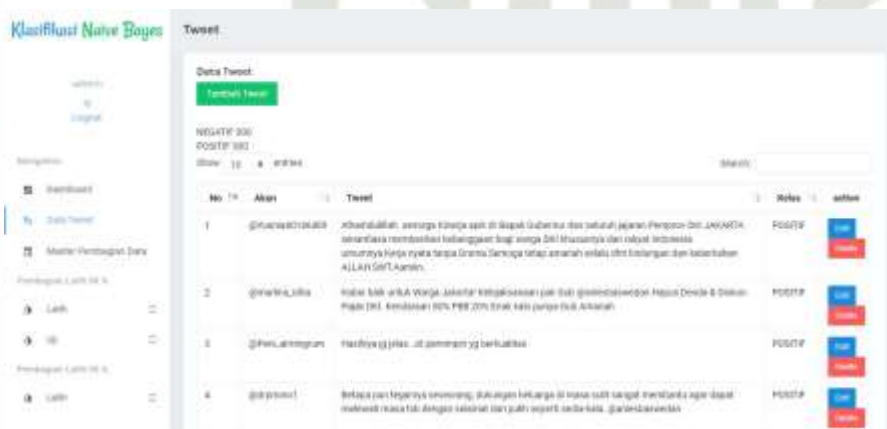
Halaman *Dashboard* memuat informasi tentang judul penelitian dan nama penelitian. Adapun di sisi *sidebar* terdapat menu yang dapat di akses untuk melakukan tahap taha klasifikasi. Menu-menu tersebut yaitu *dashboard*, data tweet, metode pembagian data, data uji dan data latih. Tampilan antarmuka menu *dashboard* dapat dilihat seperti gambar 5.1 berikut ini.



Gambar 5.1 Rancangan Antarmuka Halaman *Dashboard*

## 2. Implementasi Antarmuka Halaman Data Tweet

Halaman ini berisi data latih informasi seluruh komentar yang datanya sudah di *input* ke dalam sistem. Tampilan antarmuka data tweet dapat dilihat seperti pada gambar 5.2 berikut ini.



Gambar 5.2 Implementasi Antarmuka Halaman Data Tweet

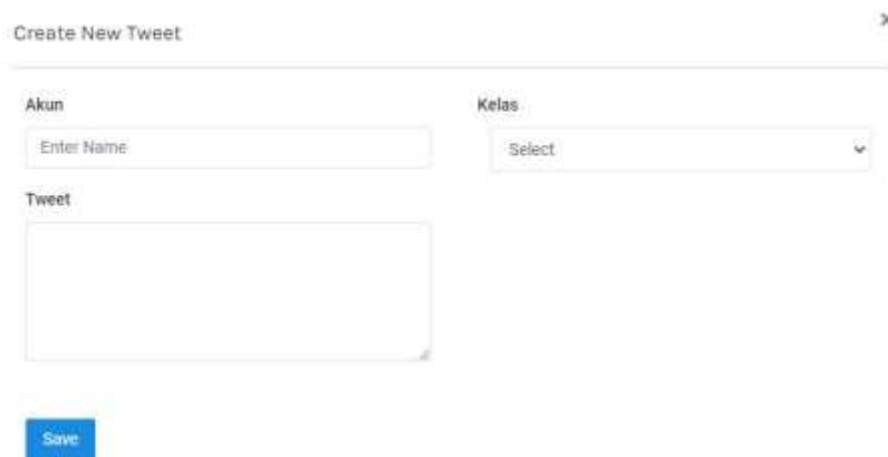


#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3. Implementasi Antarmuka Halaman Tambah Data Tweet

Halaman ini memuat tampilan yang digunakan untuk menambah data tweet, menu – menunya terdiri dari nama akun, kelas, dan tweet. Tampilan antarmuka data tweet dapat dilihat seperti pada gambar 5.3 berikut ini.



Gambar 5.3 Implementasi Halaman Tambah Data Tweet

### 4. Implementasi Halaman Hapus Data Tweet

Data tweet yang sudah di *input* ke dalam sistem dapat dirubah ataupun di hapus. Halaman *edit* memungkinkan sistem untuk merubah data. Tampilan antarmuka halaman *edit* dapat dilihat seperti pada gambar 5.4 berikut ini.



Gambar 5.4 Implementasi Halaman Hapus Data Tweet

## 5. Implementasi Antarmuka Halaman Edit Data Tweet

Halaman *edit* data tweet memungkinkan sistem untuk merubah data tweet yang sudah di *input* kedalam sistem. Tampilan antarmuka halaman *edit* dapat dilihat seperti pada gambar 5.5 berikut ini.

Gambar 5.5 Implementasi Antarmuka Halaman Edit Data Tweet

## 6. Implementasi Antarmuka Halaman Master Pembagian Data

Halaman ini terdapat tampilan tabel data sebagai data uji dengan label kelas komentarnya. Pada halaman ini tidak ada batasan untuk membagi data uji yang akan digunakan. Proses data latih 70 yang berarti apabila pengguna memasukkan angka 70 maka pengguna akan melakukan proses latih untuk 70% data, begitu juga dengan proses data latih 80 dan proses data latih 90. Tampilan antarmuka halaman data latih dapat dilihat pada gambar pada gambar 5.6 berikut ini.

No	Latih (%)	Uji (%)	Latih Positif	Uji Positif	Latih Negatif	Uji Negatif	Action
1	70	30	210	90	210	90	[Edit] [Delete]
2	80	20	240	60	240	60	[Edit] [Delete]
3	90	10	270	30	270	30	[Edit] [Delete]

Gambar 5.6 Implementasi Antarmuka Halaman Master Pembagian Data

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 5.2. Pengujian

Pengujian adalah tahapan yang dilakukan untuk melihat sejauh mana aplikasi yang dibuat ini dapat berjalan sesuai dengan analisa perancangan yang diinginkan. Selain itu pengujian dilakukan untuk mendapatkan hasil akhir yang sesuai sebagai solusi permasalahan dari penelitian ini. Pengujian dilakukan dengan dua cara yaitu *white box* dan *confusion matrix*.

### 5.2.1 Pengujian White Box

Pengujian *white box* adalah salah satu cara untuk menguji suatu aplikasi dengan cara melihat modul untuk menganalisa kode program yang telah dibuat sesuai atau tidak. Jika modul mendapatkan hasil yang tidak sesuai dengan yang diharapkan, maka dilakukan cek pengulangan pada kode program hingga sesuai dengan hasil yang diharapkan. Proses pada *white box* menggunakan data latih yang sama dengan tabel 4.2 dan pengujian dilakukan dengan data uji yang terlihat pada tabel 5.1 berikut ini:

1. Pengujian *white box* untuk Prioritas

**Tabel 5.1 Pengujian White Box untuk Prioritas**

Hasil White Box	Hasil Perhitungan Manual	Ket
Array ( [pos] => 0.33333333333333 [neg] => 0.33333333333333 )	Pos = $\frac{1}{2} = 0.33333333333333$  Neg = $\frac{1}{2} = 0.33333333333333$	Perhitungan dijadikan untuk nilai probabilitas pada setiap kategori(prioritas) yang menggunakan persamaan 2.1



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Pengujian *White Box* Frekuensi

**Tabel 5.2 Pengujian White Box untuk Frekuensi**

Hasil White Box	Hasil Perhitungan Manual	Ket
Array		
[moga] =>	moga	perhitungan untuk mencari peluang kemunculan kata
(	pos = 1	
[pos] => 1	neg = 0	
[neg] => 0		
)		
[jadi] =>	jadi	
(	pos = 1	
[pos] => 1		
[neg] => 0	neg = 0	
)		
[presiden] =>	presiden	
(	pos = 1	
[pos] => 1	neg = 0	
[neg] => 0		
)		
[republik] =>	republik	
(		

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

<pre>[pos] =&gt; 1 [neg] =&gt; 0 )  [indonesia] =&gt; (   [pos] =&gt; 1   [neg] =&gt; 0 )  [amin] =&gt; (   [pos] =&gt; 2   [neg] =&gt; 0 )</pre>	<pre>pos = 1 neg = 0  indonesia pos = 1 neg = 0  amin pos = 2 neg = 0</pre>	
Hasil White Box	Hasil Perhitungan Manual	Ket
<pre>[jangan] =&gt; (   [pos] =&gt; 0   [neg] =&gt; 1 )  [banyak] =&gt;</pre>	<pre>jangan pos = 0 neg = 1  banyak</pre>	

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

( [pos] => 0 [neg] => 1 )	pos = 0  neg = 1	
[omong] => ( [pos] => 0 [neg] => 1 )	omong  pos = 0  neg = 1	
[suruh] => ( [pos] => 0 [neg] => 1 )	suruh  pos = 0  neg = 1	
[ketua] => ( [pos] => 0 [neg] => 1 )	ketua  pos = 0  neg = 1	
[berani] => (	berani	



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

[pos] => 0	pos = 0	
[neg] => 1	neg = 1	
)		
[tidak] =>		
(		
[pos] => 0		
[net] => 0		
[neg] => 1		
)		
	tidak	
	pos = 0	
	neg = 1	

### 3. Pengujian *white box* untuk *document frequency*

**Tabel 5.3 Pengujian White Box untuk *document frequency***

Hasil White Boxes	Hasil Perhitungan Manual	Ket
Array ( [moga] => 1 [jadi] => 1 [presiden] => 1 [republik] => 1 [indonesia] => 1	moga = 1 jadi = 1 presiden = 1 republik = 1 indonesia = 1	perhitungan untuk mencari berapa banyak kemunculan kata pada setiap data

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

[amin] => 1	amin = 1	
[jangan] => 1	[jangan] => 1	
[banyak] => 1	[banyak] => 1	
[omong] => 1	[omong] => 1	
[suruh] => 1	[suruh] => 1	
[ketua] => 1	[ketua] => 1	
[berani] => 1	[berani] => 1	
[tidak] => 1	[tidak] => 1	

4. Pengujian *white box* untuk *inverse document frequency*

**Tabel 5.4 Pengujian White Box untuk *inverse document frequency***

Hasil White Boxs	Hasil Perhitungan Manual	Ket
Array ( [moga] => 0.301029996 [jadi] => 0.301029996 [presiden] => 0.301029996 [republik] => 0.301029996 [indonesia] => 0.301029996 [amin] => 0.602059991 [jangan] => 0	moga= 0.301029996 jadi = 0.301029996 presiden = 0.301029996 republik = 0.301029996 indonesia = 0.301029996 amin = 0.602059991 [jangan] => 0	hasil white box dengan perhitungan manual memiliki nilai yang sama

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

[banyak] => 0	[banyak] => 0	
[omong] => 0	[omong] => 0	
[suruh] => 0	[suruh] => 0	
[ketua] => 0	[ketua] => 0	
[berani] => 0	[berani] => 0	
[tidak] => 0	[tidak] => 0	

#### 5. Pengujian *white box* untuk *term frequency inverse document frequency*

**Tabel 5.5 Pengujian White Box untuk *term frequency inverse document frequency***

Hasil White Boxs	Hasil Perhitungan Manual	Ket
Array ( [moga] => 0.301029996 [jadi] => 0.301029996 [presiden] => 0.301029996 [republik] => 0.301029996 [indonesia] => 0.301029996 [amin] => 0.602059991 [jangan] => 0 [banyak] => 0 [omong] => 0 [suruh] => 0	Data 1  moga= 0.301029996 jadi = 0.301029996 presiden = 0.301029996 republik = 0.301029996 indonesia = 0.301029996 amin = 0.602059991 jangan => 0 banyak => 0 omong => 0 suruh => 0	memiliki nilai yang sama



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau  
State Islamic University of Sultan Syarif Kasim Riau

[ketua] => 0 [berani] => 0 [tidak] => 0	ketua => 0 berani => 0 tidak => 0	
Array ( [moga] => 0 [jadi] => 0 [presiden] => 0 [republik] => 0 [indonesia] => 0 [amin] => 0 [jangan] => 0 [banyak] => 0 [omong] => 0 [suruh] => 0 [ketua] => 0 [berani] => 0 [tidak] => 0	Data 2 moga= 0 jadi = 0 presiden = 0 republik = 0 indonesia = 0 amin = 0 jangan => 0.301029996 banyak => 0.301029996 omong => 0.301029996 suruh => 0.301029996 ketua => 0.301029996 berani => 0.301029996 tidak => 0.301029996	

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

#### 6. Pengujian *white box* untuk probabilitas

**Tabel 5.6 Pengujian White Box untuk Probabilitas**

Hasil White Boxes	Hasil Perhitungan Manual	Ket
Array ( [moga] => Array ( [pos] => 0.083014375 [neg] => 0.055555556 [jadi] => Array ( [pos] => 0.08206229 [neg] => 0.055555556 [neg] => 0.07692308 ) [presiden] => Array ( [pos] => 0.08206229 [neg] => 0.055555556	moga  pos = 0.068475263  neg = 0.055555556  jadi  pos = 0.068475263 neg = 0.055555556  presiden  pos = 0.08206229 neg = 0.055555556  republik  pos = 0.08206229 neg = 0.055555556  indonesia  pos = 0.08206229 neg = 0.055555556	perhitungan untuk nilai probabilitas setiap term dari semua dokumen data latih

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau  
State Islamic University of Sultan Syarif Kasim Riau

<pre> [net] =&gt; 0.07692308 )  [republik] =&gt; Array (   [pos] =&gt; 0.08206229   [neg] =&gt; 0.05555556   [indonesia] =&gt; Array   (     [pos] =&gt; 0.08206229     [neg] =&gt; 0.05555556   [amin] =&gt; Array   (     [pos] =&gt; 0.08206229     [neg] =&gt; 0.05555556   [jangan] =&gt; Array   (     [pos] =&gt; 0.0625     [neg] =&gt; </pre>	<pre> amin   pos =0.08206229   neg = 0.05555556  jangan   pos =0.052631579   neg = 0.0650515  banyak   pos =0.05555556   neg = 0.08206229  omong   pos =0.05555556   neg = 0.08206229  suruh   pos =0.05555556   neg = 0.08206229  ketua   pos =0.05555556   neg = 0.08206229 </pre>	
--	--	--



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

0.0650515 [banyak] => Array ( [pos] => 0.0625 [neg] => 0.0650515 [omong] => Array ( [pos] => 0.0625 [neg] => 0.0650515 [suruh] => Array ( [pos] => 0.0625 [neg] => 0.0650515 [ketua] => Array ( [pos] => 0.0625 [neg] => 0.0650515 [berani] => Array	berani  pos =0.05555556 neg = 0.08206229  tidak  pos =0.05555556 neg = 0.08206229	
---	---	--

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

( [pos] => 0.0625 [neg] => 0.0650515 [tidak] => Array ( [pos] => 0.0625 [neg] => 0.0650515		
--	--	--

7. Pengujian *White Box* untuk Probabilitas Uji

**Tabel 5.7 Pengujian *White Box* untuk Probabilitas Uji**

Hasil White Box	Hasil Perhitungan Manual	Ket
Array ( [Pos] => 5,62001	Pos = 5,62001	perhitungan untuk nilai probabilitas setiap term dari semua dokumen data uji dengan

**Tabel 5.8 Pembagian data**

No	Latih (%)	Uji (%)	Latih Positif	Uji Positif	Latih Negatif	Uji Negatif
1	70	30	210	90	210	90
2	80	20	240	60	240	60
3	90	10	270	30	270	30

### 1. Pengujian 90% dan 10%

Average			
No	Class	Nilai Rata-rata	%
1	precision	0.65428571428571	65%
2	recall	0.65	65%
3	f1 score	0.64755244755245	65%

Recall		
No	Class	Nilai Recall
1	NEGATIF	0.56666666666667
2	POSITIF	0.73333333333333

Precision		
No	Class	Nilai Precision
1	NEGATIF	0.68
2	POSITIF	0.62857142857143



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© H

#### F1 Score

No	Class	Nilai F1 Score
1	NEGATIF	0.61818181818182
2	POSITIF	0.57692307692308

#### Support

No	Class	Nilai Support
1	NEGATIF	30
2	POSITIF	30

Pada pengujian ini menggunakan 270 data latih dan 30 data uji, maka diperoleh akurasi rata-rata sebesar = 65%

2. Pengujian 80% dan 20%

#### Average

No	Class	Nilai Rate-rata	%
1	precision	0.70110754881959	70%
2	recall	0.69166666666667	69%
3	f1 score	0.68800505937742	69%

#### Recall

No	Class	Nilai Recall
1	NEGATIF	0.58333333333333
2	POSITIF	0.8

#### Precision

No	Class	Nilai Precision
1	NEGATIF	0.74468085106383
2	POSITIF	0.65753424657534

#### F1 Score

No	Class	Nilai F1 Score
1	NEGATIF	0.65420560747664
2	POSITIF	0.7218045112782

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

#### Support

No	Class	Nilai Support
1	NEGATIF	60
2	POSITIF	60

Pada pengujian ini menggunakan 240 data latih dan 60 data uji, maka diperoleh akurasi rata-rata sebesar = 70%

3. Pengujian 70% dan 30%

#### Average

No	Class	Nilai Rata-rata	%
1	precision	0.64354066985646	64%
2	recall	0.633333333333333	63%
3	f1 score	0.62869683257919	63%

#### Recall

No	Class	Nilai Recall
1	NEGATIF	0.5
2	POSITIF	0.766666666666667

#### Precision

No	Class	Nilai Precision
1	NEGATIF	0.681818181818181
2	POSITIF	0.60526315789474

#### F1 Score

No	Class	Nilai F1 Score
1	NEGATIF	0.57692307692308
2	POSITIF	0.67647058823529

#### Support

No	Class	Nilai Support
1	NEGATIF	90
2	POSITIF	90

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Pada pengujian ini menggunakan 210 data latih dan 90 data uji, maka diperoleh akurasi rata-rata sebesar = 64%

### 5.2.3 Hasil Pengujian

Berdasarkan tahapan pengujian yang telah dilakukan pada tahapan sebelumnya, maka dapat diambil kesimpulan sebagai berikut :

1. Dengan menggunakan pengujian *white box*, dapat disimpulkan bahwa proses aplikasi berhasil berjalan sesuai dengan yang diharapkan dan adanya kesesuaian dari analisa dan perancangan
2. Berdasarkan pengujian pada masing-masing dataset menggunakan *confusion matrix* menunjukkan bahwa :
  - a. Pengujian 90% dan 10% menggunakan 270 data latih dan 30 data uji, maka diperoleh akurasi rata-rata sebesar 65%
  - b. Pengujian 80% dan 20% menggunakan 240 data latih dan 60 data uji, maka diperoleh akurasi rata-rata sebesar 70%
  - c. Pengujian 70% dan 30% menggunakan 210 data latih dan 90 data uji, maka diperoleh akurasi rata-rata sebesar 64%
3. Dari pengujian menggunakan *confusion matrix* tersebut dapat diperoleh hasil tingkat akurasi tertinggi sebesar 70% yaitu pada perbandingan 80:20 dan akurasi terendah sebesar 64% pada perbandingan 70:30.



## BAB VI

### PENUTUP

#### 6.1. Kesimpulan

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan, maka dapat ditarik beberapa kesimpulan sebagai berikut :

1. Metode *Naive Bayes Classifier* dapat diterapkan dengan baik untuk melakukan klasifikasi sentimen pada klasifikasi sentimen terhadap Gubernur DKI Jakarta di media sosial *twitter* menggunakan metode *Naive Bayes classifier*.
2. Metode *Naive Bayes Classifier* mudah untuk di terapkan dan memiliki hasil yang bagus.
3. Tingkat akurasi mencapai nilai tertinggi yaitu 70% pada perbandingan 80:20.

#### 4. Saran

Adapun saran yang dapat diberikan untuk penelitian lanjutanyang terkait dengan penelitian ini adalah :

1. Aplikasi klasifikasi komentar ini bisa dikembangkan dengan menggunakan metode-metode klasifikasi lain.
2. Aplikasi ini dapat dikembangkan untuk platform social media lainnya seperti facebook, instagram, path, kaskus, dan lainnya.

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
- ## DAFTAR PUSTAKA
- Ahmad Fathan, Azhari SN. 2014. "Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter."
- Arifin, A. Z., Mahendra, I. P., & Ciptaningtyas, H. T. 2009. "Penggunaan Algoritma Semut Dan Confix Stripping Stemmer Untuk Klasifikasi Dokumen Berita Berbahasa Indonesia." *jurnali lmiah*.
- Attanti, Dian Putri, Abdul Syukur, Adi Prihandono, and De Rosal I Moses Setiadi. 2018. "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes." : 8–9.
- B.J.M.Putra,A,A.Helen, A.R.Barakbah. 2018. 6 Emit. Int. J. Eng. Technol *Rule-Based Sentiment Degree Measurement of Opinion Mining of Community Participatory in the Government of Surabaya*.
- DataBooks, Kata Data. 2017. "Salip Facebook, Twitter Paling Aktif Di Indonesia." *DataBooks, Kata Data*: 2020.
- Hamzah, Amir. 2012. "Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita Dan Abstract Akademis." *Institut Sains dan Teknologi AKPRIND*.
- Handayani, Fitri, and Feddy Setio Pribadi. 2015. "Implementasi Algoritma Naive Bayes Classifier Dalam Pengklasifikasian Teks Otomatis Pengaduan Dan Pelaporan Masyarakat Melalui Layanan Call Center 110." *Jurnal Teknik Elektro* 7(1): 19–24.
- Hendy Evan, F., & Sigit Purnomo, Y. W. 2014. "Pembangunan Perangkat Lunak Peringkat Dokumen Dari Banyak Sumber Menggunakan Sentence Scoring Dengan Metode TF-IDF." *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, 21(17), 1907–5022.
- Jayanti, Linda, Steven Ray Sentinuwo, Oktavian A Lantang, and Agustinus Jacobus. 2016. "Analisa Pola Penyalahgunaan Facebook Sebagai Alat Kejahatan Trafficking Menggunakan Data Mining." *E-journal Teknik Informatika* 8(1): 30–35.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.
- L Novianti, A. Ardiyanti dan A. Prima. 2012. "Pengklasifikasian Dokumen Berita Berbahasa Indonesia Menggunakan Latent Semantic Indexing (LSI) Dan Support Vector Machine (SVM)."
- Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. 2013. "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8206 LNCS: 194–201.
- Nugroho, Moh Aziz. 2016. "Klasifikasi Dokumen Komentar Pada Situs Twitter Menggunakan Algoritma K-Nearest Neighbor (K-Nn)." *Universitas Dian Nuswantoro* (5).
- Prasetyo, E. 2012. *Data Mining : Konsep Dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi Offset.
- Rustiana, Deden, and Nina Rahayu. 2017. "Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes." 8(1): 113–20.
- Sanjaya dan Absar. 2015. "Pengelompokkan Dokumen Menggunakan Winnowing Fingerprint Dengan Metode K-Nearest Neighbour." *Jurnal CoreIT* Vol. 1: No. 2.,
- Tanesab, Fiktor Imanuel, Irwan Sembiring, and Hindriyanto Dwi Purnomo. 2017. "Sentiment Analysis Model Based On Twitter Comment Using Support Vector Machine." *International Journal of Computer Science and Software Engineering* ISSN 6(8): 2409–4285.
- Yong-Feng, S. & Yan-Ping, Z. 2004. *Comparison of text categorization algorithms*. Wuhan University Journal of Natural Sciences, 9, 798-804.
- Widodo, A. W., Mahmudy, W. F. & Maisuroh, M. 2007. *Klasifikasi artikel otomatis, sebuah kajian eksperimen. FKP2T*, 2.
- Miao, Y.-Q. & Kamel, M. 2011. *Pairwise optimized Rocchio algorithm for text categorization. Pattern Recognition Letters*, 32, 375-382.